CORRECTED VERSION*

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C07K 14/705, 16/28, C12N 15/09, 15/63, 15/70, 15/79, 15/80, C12Q 1/00, 1/68

A2

(11) International Publication Number: WO 99/52945

(43) International Publication Date: 21 October 1999 (21.10.99)

(21) International Application Number: PCT/US99/08395

(22) International Filing Date: 16 April 1999 (16.04.99)

(30) Priority Data:

09/061,753	16 April 1998 (16.04.98)	US
09/	16 April 1999 (16.04.99)	US

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With declaration under Article 17(2)(a); without abstract; title not checked by the International Searching Authority.

(71) Applicant: MILLENIUM PHARMACEUTICALS, INC. [US/US]; 640 Memorial Drive, Cambridge, MA 02139 (US).

(72) Inventors: GONZALO, Jose, Angel; Apartment 102, 285 Harvard Street, Cambridge, MA 02139 (US). GUTIERREZ-RAMOS, Jose, Carlos; 31 Littles Point Road, Swampscott, MA 01902 (US).

(74) Agent: MEIKLEJOHN, Anita, L.; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).

(54) Title: NOVEL MOLECULES OF THE BGCKr-RELATED PROTEIN FAMILY AND USES THEREOF

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

NOVEL MOLECULES OF THE BGCKr-RELATED PROTEIN FAMILY AND USES THEREOF

Background of the Invention

The G-protein coupled receptors form a superfamily of receptors that share several common structural features including an extracellular aminoterminus, seven transmembrane spanning domains and a cytoplasmic carboxyterminus with clustered serine and threonine residues. The sequences of more than 200 G-protein coupled receptors have been reported, and include receptors for neurotransmitters, hormones, chemoattractants, odorants, and light (Probst et al. (1992) *DNA Cell Biol*, 11:1-20).

G-protein coupled receptors normally transduce their signals through a heterotrimeric GTP-binding protein (Horuk (1994) *TIPS*, 15:159-165; Rodbell, (1997) *Advan. Enzyme Regul.* 37:427-435). This signalling mechanism has been especially well characterized in the hormone responsive B_2-adrenergic receptor, which mediates catecholamine stimulation of adenyl cyclase, and the light receptor rhodopsin, which mediates phototransduction in retinal rod cells. Following ligand binding, the G-protein coupled receptor undergoes an alteration in its conformation so as to enable it to activate the G protein in its lipid bilayer. The G protein then undergoes a reaction cycle which involves guanine nucleotide binding and hydrolysis, G protein subunit dissociation, and interactions with an effector molecule such as adenyl cyclase. This receptor stimulation is tightly regulated such that once the receptor interacts with its ligand, the receptor undergoes rapid and reversible loss of responsiveness to subsequent stimulation. Multiple mechanisms can result

in G-protein receptor desensitization including phosphorylation and internalization of the receptor.

Leukocytes respond to chemokines through G protein-coupled receptors. Chemokines constitute a large family of small (8-10 kDa) inducible, secreted cytokines, which are produced by a variety of cell types (Baggiolini et al. (1997) *Ann. Rev. Immunol.* 15:675-705), and play a crucial role in the recruitment and activation of leukocytes at a site of inflammation. Members of the chemokine family can be subdivided into two families, the CXC and the CC family. The CXC family contain a single amino acid residue between the first and second cysteine residue whereas those of the CC family have adjacent cysteine residues.

Chemokine receptors form a distinct group of structurally related proteins within the superfamily of G-protein receptors. These receptors have a number of conserved structural motifs which can usually be found in their transmembrane domains. In addition, they contain 2 conserved cysteines, one at the amino terminal domain and the other in the third extracellular loop, which are thought to form a disulfide bond critical for the formation of the ligand binding pocket ((Baggiolini et al. *supra*). Based on their overall sequence identity, chemokine receptors can be divided into two subgroups, the CXC chemokine receptors which include IL-8RA and IL-8RB (Holmes et al. (1991) *Science* 253:1278-80, 1991; Murphy et al. (1991) *Science* 253:1278-80), and the CC chemokine receptors which include CKR-1, MCP-1Ra and MCP-1Rb (Neote et al. (1993) *Cell* 72:415-425; Gao et al. (1993) *J. Exp. Med.* 177:1421-1427, and Charo et al. (1994) *Proc. Natl. Acad. Sci. USA*, 91:2752-2756).

Summary of the Invention

The present invention is based, at least in part, on the discovery of a gene encoding BGCKr, a G-protein coupled receptor that is predicted to be a member of the chemokine receptor superfamily. The BGCKr cDNA described below (SEQ ID NO:4) has a 1050 nucleotide open reading frame (nucleotides 70-1119 of SEQ ID NO:4; SEQ ID NO:6) which encodes a 350 amino acid protein (SEQ ID NO:5).

The amino acid sequence of BGCKr (corresponding to amino acid 1 to amino acid 350 of SEQ ID NO:5) was aligned with the amino acid sequences of bovine gustatory receptor (SEQ ID NO:10, accesssion number P35350), ceraeBonzo (SEQ ID NO:11, accession number 018983), human CCR6 (SEQ ID NO:12, accession number P51684), human CCR7 (SEQ ID NO:13, accession number P32248), human GPR-9-6 (SEQ ID NO:14, accession number P51686), murine CCR6 (SEQ ID NO:15, accession number AJ222714), and murine CCR7 (SEQ ID NO:16, accession number W48723) as shown in Figure 5.

The BGCKr protein possesses seven predicted transmembrane (TM) domains. The first TM domain extends from amino acid 42 or 44 to amino acid 66 of SEQ ID NO:5; the second TM domain extends from amino acid 78 or 79 to amino acid 98 or 99 of SEQ ID NO:5; the third TM domain extends form amino acid 114 to amino acid 133 or 135 of SEQ ID NO:5; the fourth TM domain extends from amino acid 154 or 156 to amino acid 175 or 176 of SEQ ID NO:5; the fifth TM domain extends from amino acid 198 or 202 to amino 218 or 224 of SEQ ID NO:5; the sixth TM domain extends from amino acid 238 or 241 to amino acid 259 or 262 of SEQ ID NO:5; and the seventh TM domain extends from amino acid 287 or 291 to amino acid 307 or 310 of SEQ ID NO:5.

Because chemokines play a role in recruitment and activation of leukocytes, one can, by altering the

expression or activity of chemokine receptors such as BGCKr, modulate inflammatory processes and the chemoattractant activity of leukocytes. Moreover, since chemokines may play a role in angiogenesis, proliferation, tumor growth, allergic reactions and the entry of human immunodeficiency virus (HIV) into cells, it may be possible to modulate these processes by altering the expression or activity of BGCKr and thus treat cancer, allergic reactions, and HIV infections.

The BGCKr molecules of the present invention are useful as modulating agents, as diagnostic agents, and targets for drug screening. Accordingly, in one aspect, this invention provides isolated nucleic acid molecules encoding BGCKr proteins or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of BGCKr-encoding nucleic acids.

The invention features a nucleic acid molecule which is at least 45% (or 55%, 65%, 75%, 85%, 95%, or 98%) identical to the nucleotide sequence shown in SEQ ID NO:1, 3, 4, 6, or 7, or the nucleotide sequence of the cDNA insert of the plasmid deposited with ATCC as Accession Number 209756 (the "cDNA of ATCC 209756"), or a complement thereof. The invention also features a nucleic acid molecule which includes a nucleotide sequence encoding a protein having an amino acid sequence that is at least 45% (or 55%, 65%, 75%, 85%, 95%, or 98%) identical to the amino acid sequence of SEQ ID NO:2, 5, or 8, encoded by the cDNA of ATCC 209756. In preferred embodiments, a BGCKr nucleic acid molecule has the nucleotide sequence shown SEQ ID NO:1, 3, 4, 6, 7, or the nucleotide sequence of the cDNA of ATCC 209756.

Also within the invention is an isolated BGCKr protein having an amino acid sequence that is at least

about 65%, preferably 75%, 85%, 95%, or 98% identical to the BGCKr amino acid sequence of SEQ ID NO:2, 5, or 8.

Also within the invention are: an isolated BGCKr protein which is encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 65%, preferably 75%, 85%, or 95% identical to SEQ ID NO:3, 6, or the cDNA of ATCC 209756; an isolated BGCKr protein which is encoded by a nucleic acid molecule having a nucleotide sequence at least about 65% preferably 75%, 85%, or 95% identical to SEQ ID NO:1, 4, or 7; and an isolated BGCKr protein which is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:3, 6, or the non-coding strand of the cDNA of ATCC 209756. Preferably, the nucleic acid molecule encodes a naturally-occurring transmembrane protein having seven transmembrane domains. The nucleic acid molecule preferably has the sequence of a naturally-occurring nucleic acid molecule or a naturally-occurring nucleic acid molecule in which "U" is replaced by "T."

Another embodiment of the invention features BGCKr nucleic acid molecules which specifically detect BGCKr nucleic acid molecules relative to nucleic acid molecules encoding other members of the chemokine receptor superfamily. For example, in one embodiment, a BGCKr nucleic acid molecule hybridizes under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, or the cDNA of ATCC 209756, or a complement thereof. In another embodiment, the BGCKr nucleic acid molecule is at least 300 (400, 450, 500, 550, 600, 800, 900, 1000, 1050, 1100, 1119) nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence shown in SEQ ID NO:1,

3, 4, 6, 7, the cDNA of ATCC 209756, or a complement thereof. In a preferred embodiment, an isolated BGCKr nucleic acid molecule comprises SEQ ID NO:1, 3, 4, 6, 7, or a complement thereof. In another embodiment the invention provides an isolated nucleic acid molecule which is antisense to the coding strand of a BGCKr nucleic acid.

Another aspect of the invention provides a vector, e.g., a recombinant expression vector, comprising a BGCKr nucleic acid molecule of the invention. In another embodiment the invention provides a host cell containing such a vector. The invention also provides a method for producing BGCKr protein by culturing, in a suitable medium, a host cell of the invention containing a recombinant expression vector such that a BGCKr protein is produced.

Another aspect of this invention features isolated or recombinant BGCKr proteins and polypeptides. In one embodiment, an isolated BGCKr protein is soluble or secreted or retained in an intracellular compartment and lacks a transmembrane or cytoplasmic domain. In other embodiments, a BGCKr protein possesses at least one biological activity possessed by naturally occurring human BGCKr.

The BGCKr proteins of the present invention, or biologically active portions thereof, can be operatively linked to a non-BGCKr polypeptide (e.g., heterologous amino acid sequences) to form BGCKr fusion proteins. The invention further features antibodies that specifically bind BGCKr proteins, such as monoclonal or polyclonal antibodies. In addition, the BGCKr proteins or biologically active portions thereof can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

In another aspect, the present invention provides a method for detecting the presence of BGCKr activity or expression in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of BGCKr activity such that the presence of BGCKr activity is detected in the biological sample.

In another aspect, the invention provides a method for modulating BGCKr activity comprising contacting a cell with an agent that modulates (inhibits or stimulates) BGCKr activity or expression such that BGCKr activity or expression in the cell is modulated. In one embodiment, the agent is an antibody that specifically binds to BGCKr protein. In another embodiment, the agent modulates expression of BGCKr by modulating transcription of a BGCKr gene, splicing of a BGCKr mRNA, or translation of a BGCKr mRNA. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of the BGCKr mRNA or the BGCKr gene.

In one embodiment, the methods of the present invention are used to treat a subject having a disorder characterized by aberrant BGCKr protein or nucleic acid expression or activity by administering an agent which is a BGCKr modulator to the subject. In one embodiment, the BGCKr modulator is a BGCKr protein. In another embodiment the BGCKr modulator is a BGCKr nucleic acid molecule. In other embodiments, the BGCKr modulator is a peptide, peptidomimetic, or other small molecule. In a preferred embodiment, the disorder characterized by aberrant BGCKr protein or nucleic acid expression is a inflammatory disorder.

The present invention also provides a diagnostic assay for identifying the presence or absence of a genetic lesion or mutation characterized by at least one of: (i) aberrant modification or mutation of a gene

encoding a BGCKr protein; (ii) mis-regulation of a gene encoding a BGCKr protein; and (iii) aberrant post-translational modification of a BGCKr protein, wherein a wild-type form of the gene encodes an protein with a BGCKr activity.

In another aspect, the invention provides a method for identifying a compound that binds to or modulates the activity of a BGCKr protein. In general, such methods entail measuring a biological activity or a BGCKr protein in the presence and absence of a test compound, and identifing those compounds which alter the activity of the BGCKr protein.

The invention also features methods for identifying a compound which modulates the expression of BGCKr by measuring the expression of BGCKr in the presence and absence of a compound.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figure 1 depicts the cDNA sequence (SEQ ID NO:1) and predicted amino acid sequence (SEQ ID NO:2) of a partial human BGCKr clone. The open reading frame of SEQ ID NO:1 extends from nucleotide 19 to nucleotide 1017 (SEQ ID NO:3).

Figure 2 depicts the cDNA sequence (SEQ ID NO:4) and predicted amino acid sequence (SEQ ID NO:5) of full-length human BGCKr. The open reading frame of SEQ ID NO:4 extends from nucleotide 70 to nucleotide 1119 of SEQ ID NO:4 (SEQ ID NO:6).

Figure 3 depicts the cDNA sequence (SEQ ID NO:7) and predicted amino acid sequence (SEQ ID NO:8) of murine BGCKr. The open reading frame of SEQ ID NO:7 extends from nucleotide 1 to nucleotde 1050 of SEQ ID NO:7 (SEQ ID NO:7).

Figure 4 depicts a hydropathy plot of human BGCKr. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by vertical bars.

Figure 5 depicts an alignment of the amino acid sequences of full-length human BGCKr, bovine gustatory receptor (SEQ ID NO:10, accesssion number P35350), ceraeBonzo (SEQ ID NO:11, accession number 018983), human CCR6 (SEQ ID NO:12, accession number P51684), human CCR7 (SEQ ID NO:13, accession number P32248), human GPR-9-6 (SEQ ID NO:14, accession number P51686), murine CCR6 (SEQ ID NO:15, accession number AJ222714), and murine CCR7 (SEQ ID NO:16, accession number W48723). A consenus sequence is also depicted (SEQ ID NO:20).

Figure 6 depicts an alignment of the amino acid sequences of murine BGCKr (SEQ ID NO:8), human BGCKr (SEQ ID NO:5), and bovine gustatory receptor (SEQ ID NO:10, accession number P35350). A consensus sequence is also depicted (SEQ ID NO:21).

Figure 7 depicts the results of an analysis of BGCKr expression in tissues isolated from Balb/C mice.

Figure 8 depicts the results of an analysis of BGCKr expression in cells isolated from Balb/C mice.

Figure 9 depicts the results of an analysis of BGCKr expression in various human tissues, isolated human cells, and human cell lines.

Figure 10 depicts an alignment of amino acids 58-303 of human BGCKr (SEQ ID NO:19) with a seven transmembrane receptor consensus sequence derived from a Hidden Markov Model (PF00001).

Detailed Description of the Invention

The present invention is based, in part, on the discovery of a cDNA molecule encoding human BGCKr, a member of the chemokine receptor superfamily.

A nucleotide sequence encoding a human BGCKr protein is shown in Figure 2 (SEQ ID NO:4; SEQ ID NO:6 includes the open reading frame only). A predicted amino acid sequence of BGCKr protein is also shown in Figure 2 (SEQ ID NO:5).

The human BGCKr cDNA of Figure 2 (SEQ ID NO:4), which is approximately 1232 nucleotides long including untranslated regions, encodes a protein amino acid having a molecular weight of approximately 39.9 kDa (excluding post-translational modifications). A plasmid containing a BGCKr cDNA was deposited with American Type Culture Collection (ATCC), Manassas, Virginia on April 8, 1998, and assigned Accession Number 209756. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Human BGCKr is one member of a family of molecules (the "chemokine receptor family") having certain conserved structural and functional features. The term "family" when referring to the protein and nucleic acid molecules of the invention is intended to mean two or more proteins or nucleic acid molecules having a common structural domain and having sufficient amino acid or nucleotide sequence identity as defined herein. Such family members can be naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin and a homologue of that protein of murine origin (e.g., the murine BGCKr of Figure 3), as well as a second, distinct

protein of human origin and a murine homologue of that protein. Members of a family may also have common functional characteristics.

Preferred BGCKr polypeptides of the present invention have an amino acid sequence sufficiently identical to the sequence of SEQ ID NO:5. As used herein, the term "sufficiently identical" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (e.g., an amino acid residue which has a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences have a common structural domain and/or common functional activity. For example, amino acid or nucleotide sequences which contain a common structural domain having about 65% identity, preferably 75% identity, more preferably 85%, 95%, or 98% identity are defined herein as sufficiently identical.

Also within the invention is a BGCKr polypeptide lacking a signal sequence (e.g., a polypeptide having the sequence of SEQ ID NO:5 from amino acid 18-350, 19-350, 19-350, 20-350, 21-350, 22-350, 23-350, 24-350, 35-350, 26-350, 27-350, 28-350, 29-350, or 30-350).

Also within the invention are polypeptides comprising one or more of the following portions of SEQ ID NO:5: amino acids 42-66, 78-99, 114-135, 154-176, 202-224, 241-259, all of which correspond to predicted transmembrane domains.

As used interchangeably herein a "BGCKr activity", "biological activity of BGCKr" or "functional activity of BGCKr", refers to an activity exerted by a BGCKr protein, polypeptide or nucleic acid molecule on a BGCKr responsive cell as determined *in vivo*, or *in vitro*, according to standard techniques. A BGCKr activity can be a direct activity, such as an association with or an

enzymatic activity on a second protein or an indirect activity, such as a cellular signaling activity mediated by interaction of the BGCKr protein with a second protein.

Accordingly, one embodiment of the invention features isolated BGCKr proteins and polypeptides having a BGCKr activity.

Another embodiment of the invention features BGCKr molecules which contain a signal sequence. Generally, a signal sequence (or signal peptide) is a peptide containing about 18 to 30 amino acids which occurs at the extreme N-terminal end of secretory and integral membrane proteins and which contains large numbers of hydrophobic amino acid residues and serves to direct a protein containing such a sequence to a lipid bilayer. Also within the invention is mature BGCKr proteins or polypeptides which lack a signal sequence.

Various aspects of the invention are described in further detail in the following subsections.

I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode BGCKr proteins or biologically active portions thereof, as well as nucleic acid molecules sufficient for use as hybridization probes to identify BGCKr-encoding nucleic acids (e.g., BGCKr mRNA) and fragments for use as PCR primers for the amplification or mutation of BGCKr nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences (preferably protein encoding sequences) which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated BGCKr nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, or the cDNA of ATCC 209756, or a complement of any of these nucleotide sequences, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or portion of the nucleic acid sequences of SEQ ID NO:1, 3, 4, 6, 7, or the cDNA of ATCC 209756 as a hybridization probe, BGCKr nucleic acid molecules can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook et al., eds., *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

A nucleic acid of the invention can be amplified using cDNA, mRNA or genomic DNA as a template and

appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to BGCKr nucleotide sequences can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1, 3, 4, 6, 7, or the cDNA of ATCC 209756, or a portion thereof. A nucleic acid molecule which is complementary to a given nucleotide sequence is one which is sufficiently complementary to the given nucleotide sequence that it can hybridize to the given nucleotide sequence thereby forming a stable duplex.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of a nucleic acid sequence encoding BGCKr, for example, a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of BGCKr. The nucleotide sequence determined from the cloning of the human BGCKr gene allows for the generation of probes and primers designed for use in identifying and/or cloning BGCKr homologues in other cell types, e.g., from other tissues, as well as BGCKr homologues from other mammals. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 50, 100, 150, 200, 250, 300, 350 or 400 consecutive nucleotides of the sense or anti-sense sequence of SEQ ID NO:1, 3, 4, 6, 7, or the

cDNA of ATCC 209756, or of a naturally occurring mutant of SEQ ID NO:1, 3, 4, 6, 7, or the cDNA of ATCC 209756.

Probes based on the human BGCKr nucleotide sequence can be used to detect transcripts or genomic sequences encoding the same or identical proteins. The probe comprises a label group attached thereto, e.g., a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a BGCKr protein, such as by measuring a level of a BGCKr-encoding nucleic acid in a sample of cells from a subject, e.g., detecting BGCKr mRNA levels or determining whether a genomic BGCKr gene has been mutated or deleted.

A nucleic acid fragment encoding a "biologically active portion of BGCKr" can be prepared by isolating a portion of SEQ ID NO:1, 3, 4, 6, 7, or the nucleotide sequence of the cDNA of ATCC 209756 which encodes a polypeptide having a BGCKr biological activity, expressing the encoded portion of BGCKr protein (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of BGCKr.

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, or the cDNA of ATCC 209756 due to degeneracy of the genetic code and thus encode the same BGCKr protein as that encoded by the nucleotide sequence shown in SEQ ID NO:1, 3, 4, 6, 7, or the cDNA of ATCC 209756.

In addition to the human BGCKr nucleotide sequence shown in SEQ ID NO:1, 3, 4, 6, 7, or the cDNA of ATCC 209756, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of BGCKr may exist within a population (e.g., the human population). Such genetic

polymorphism in the BGCKr gene may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a BGCKr protein, preferably a mammalian BGCKr protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the BGCKr gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in BGCKr that are the result of natural allelic variation and that do not alter the functional activity of BGCKr are intended to be within the scope of the invention.

The invention includes polypeptides having an amino acid sequence which differs from that of SEQ ID NO:2, 5, or 8 at 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35 or fewer amino acid residues. The amino acid residues can be replaced by conservative or non-conservative or non-conservative substitution.

Moreover, nucleic acid molecules encoding BGCKr proteins from other species (BGCKr homologues), which have a nucleotide sequence which differs from that of a human BGCKr, are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the BGCKr cDNAs of the invention can be isolated based on their identity to the human BGCKr nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. For example, a cDNA encoding a soluble human BGCKr can be isolated based on hybridization cDNA encoding a human membrane-bound BGCKr. Likewise, cDNA encoding a

membrane-bound human BGCKr can be isolated based on its hybridization to cDNA encoding to soluble human BGCKr.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15 (30, 50, 100, 250, or 500) nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence, preferably the coding sequence, of SEQ ID NO:1, 3, 4, 6, 7, or the cDNA of ATCC 209756.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% (65%, 70%, preferably 75%) identical to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:1, 3, 4, 6, 7, the cDNA of ATCC 209756 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in natúre (e.g., encodes a natural protein).

In addition to naturally-occurring allelic variants of the BGCKr sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, the cDNA of ATCC 209756, thereby leading to changes in the amino

acid sequence of the encoded BGCKr protein, without altering the functional ability of the BGCKr protein. For example, one can make nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of BGCKr (e.g., the sequence of SEQ ID NO:2) without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the BGCKr proteins of various species are predicted to be particularly unamenable to alteration.

For example, preferred BGCKr proteins of the present invention, contain at least one transmembrane domain of the receptor. Such conserved domains are less likely to be amenable to mutation. Other amino acid residues, however, (e.g., those that are not conserved or only semi-conserved among BGCKr of various species) may not be essential for activity and thus are likely to be amenable to alteration.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding BGCKr proteins that contain changes in amino acid residues that are not essential for activity. Such BGCKr proteins differ in amino acid sequence from SEQ ID NO:2, 5, or 8 yet retain biological activity. In one embodiment, the isolated nucleic acid molecule includes a nucleotide sequence encoding a protein that includes an amino acid sequence that is at least about 45% identical, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:5.

An isolated nucleic acid molecule encoding a BGCKr protein having a sequence which differs from that of SEQ ID NO:2 5, or 8 can be created by introducing one or more

nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1, 3, 6, the cDNA of ATCC 209756 such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in BGCKr is preferably replaced with another amino acid residue from the same side chain family. Alternatively, mutations can be introduced randomly along all or part of a BGCKr coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for BGCKr biological activity to identify mutants that retain activity. Following mutagenesis, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

A mutant BGCKr protein or a BGCKr polypeptide can be assayed for: (1) the ability to form protein:protein

interactions with proteins in the BGCKr signalling pathway, other cell-surface proteins, or biologically active portions thereof; (2) the ability to bind a BGCKr ligand; or (3) the ability of a mutant BGCKr protein to bind to an intracellular target protein or biologically active portion thereof.

The present invention encompasses antisense nucleic acid molecules, i.e., molecules which are complementary to a sense nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire BGCKr coding strand, or to only a portion thereof, e.g., all or part of the protein coding region (or open reading frame). An antisense nucleic acid molecule can be antisense to a noncoding region of the coding strand of a nucleotide sequence encoding BGCKr. The noncoding regions ("5' and 3' untranslated regions") are the 5' and 3' sequences which flank the coding region and are not translated into amino acids.

Given the coding strand sequences encoding BGCKr disclosed herein (e.g., SEQ ID NO:1, 3, 4, 6, and 7) antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of BGCKr mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of BGCKr mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of BGCKr mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic

acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an

antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a BGCKr protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention include direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

An antisense nucleic acid molecule of the invention can be an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which,

contrary to the usual β-units, the strands run parallel to each other (Gaultier et al. (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al. (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett.* 215:327-330).

The invention also encompasses ribozymes. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave BGCKr mRNA transcripts to thereby inhibit translation of BGCKr mRNA. A ribozyme having specificity for a BGCKr-encoding nucleic acid can be designed based upon the nucleotide sequence of a BGCKr cDNA disclosed herein (e.g., SEQ ID NO:1, 3, 4, 6). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a BGCKr-encoding mRNA. See, e.g., Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742. Alternatively, BGCKr mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel and Szostak (1993) *Science* 261:1411-1418.

The invention also encompasses nucleic acid molecules which form triple helical structures. For example, BGCKr gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the BGCKr (e.g., the BGCKr promoter and/or enhancers) to form triple helical structures that prevent transcription of the BGCKr gene in target cells.

See generally, Helene (1991) *Anticancer Drug Des.* 6(6):569-84; Helene (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

In preferred embodiments, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup et al. (1996) *Bioorganic & Medicinal Chemistry* 4 (1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup et al. (1996) *supra*; Perry-O'Keefe et al. (1996) *Proc. Natl. Acad. Sci. USA* 93: 14670-675.

PNAs of BGCKr can be used therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, e.g., inducing transcription or translation arrest or inhibiting replication. PNAs of BGCKr can also be used, e.g., in the analysis of single base pair mutations in a gene by, e.g., PNA directed PCR clamping; as 'artificial restriction enzymes when used in combination with other enzymes, e.g., S1 nucleases (Hyrup (1996) *supra*; or as probes or primers for DNA sequence and hybridization (Hyrup (1996) *supra*; Perry-O'Keefe et al. (1996) *Proc. Natl. Acad. Sci. USA* 93: 14670-675).

In another embodiment, PNAs of BGCKr can be modified, e.g., to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of BGCKr can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, e.g., RNAse H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) *supra* and Finn et al. (1996) *Nucleic Acids Research* 24(17):3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs, e.g., 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used as a between the PNA and the 5' end of DNA (Mag et al. (1989) *Nucleic Acid Res.* 17:5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn et al. (1996) *Nucleic Acids Research* 24(17):3357-63). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser et al. (1975) *Bioorganic Med. Chem. Lett.* 5:1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger et al. (1989) *Proc. Natl. Acad. Sci. USA*

86:6553-6556; Lemaitre et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (*See, e.g.*, Krol et al. (1988) *Bio/Techniques* 6:958-976) or intercalating agents (*See, e.g.*, Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, e.g., a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

II. Isolated BGCKr Proteins and Anti-BGCKr Antibodies

One aspect of the invention pertains to isolated BGCKr proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-BGCKr antibodies. In one embodiment, native BGCKr proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, BGCKr proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a BGCKr protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the BGCKr protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of BGCKr protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly

produced. Thus, BGCKr protein that is substantially free of cellular material includes preparations of BGCKr protein having less than about 30%, 20%, 10%, or 5% (by dry weight) of non-BGCKr protein (also referred to herein as a "contaminating protein"). When the BGCKr protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, 10%, or 5% of the volume of the protein preparation. When BGCKr protein is produced by chemical synthesis, it is preferably substantially free of chemical precursors or other chemicals, i.e., it is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. Accordingly such preparations of BGCKr protein have less than about 30%, 20%, 10%, 5% (by dry weight) of chemical precursors or non-BGCKr chemicals.

Biologically active portions of a BGCKr protein include peptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the BGCKr protein (e.g., the amino acid sequence shown in SEQ ID NO:2, 5, or 8), which include less amino acids than the full length BGCKr proteins, and exhibit at least one activity of a BGCKr protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the BGCKr protein. A biologically active portion of a BGCKr protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length. Preferred biologically active polypeptides include one or more identified BGCKr structural domains, e.g., a transmembrane domain of the receptor.

Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one

or more of the functional activities of a native BGCKr protein. Preferred BGCKr protein has the amino acid sequence of SEQ ID NO:2, 5, or 8. Other useful BGCKr proteins are substantially identical to SEQ ID NO:2, 5, or 8 and retain the functional activity of the protein of SEQ ID NO:2, 5, or 8 yet differ in amino acid sequence due to natural allelic variation or mutagenesis. Accordingly, a useful BGCKr protein is a protein which includes an amino acid sequence at least about 45%, preferably 55%, 65%, 75%, 85%, 95%, or 99% identical to the amino acid sequence of SEQ ID NO:2, 5, or 8 and retains the functional activity of the BGCKr proteins of SEQ ID NO:2. In other instances, the BGCKr protein is a protein having an amino acid sequence 55%, 65%, 75%, 85%, 95%, or 98% identical to the BGCKr transmembrane domain of the chemokine receptor. In a preferred embodiment, the BGCKr protein retains the functional activity of the BGCKr protein of SEQ ID NO:2.

To determine the percent identity of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = # of identical positions/total # of positions x 100).

The determination of percent homology between two sequences can be accomplished using a mathematical

algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990) *Proc. Nat'l Acad. Sci. USA* 87:2264-2268, modified as in Karlin and Altschul (1993) *Proc. Nat'l Acad. Sci. USA* 90:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, et al. (1990) *J. Mol. Biol.* 215:403-410. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to BGCKr nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to BGCKr protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) *Nucleic Acids Res.* 25:3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov. Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, CABIOS (1989). Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, only exact matches are counted.

The invention provides BGCKr chimeric or fusion proteins. As used herein, a BGCKr "chimeric protein" or

"fusion protein" comprises a BGCKr polypeptide operatively linked to a non-BGCKr polypeptide. A "BGCKr polypeptide" refers to a polypeptide having an amino acid sequence corresponding to BGCKr, whereas a "non-BGCKr polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially identical to the BGCKr protein, e.g., a protein which is different from the BGCKr protein and which is derived from the same or a different organism. Within a BGCKr fusion protein the BGCKr polypeptide can correspond to all or a portion of a BGCKr protein, preferably at least one biologically active portion of a BGCKr protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the BGCKr polypeptide and the non-BGCKr polypeptide are fused in-frame to each other. The non-BGCKr polypeptide can be fused to the N-terminus or C-terminus of the BGCKr polypeptide.

One useful fusion protein is a GST-BGCKr fusion protein in which the BGCKr sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant BGCKr.

In another embodiment, the fusion protein is a BGCKr protein containing a heterologous signal sequence at its N-terminus. For example, the native BGCKr signal sequence can be removed and replaced with a signal sequence from another protein. In certain host cells (e.g., mammalian host cells), expression and/or secretion of BGCKr can be increased through use of a heterologous signal sequence. For example, the gp67 secretory sequence of the baculovirus envelope protein can be used as a heterologous signal sequence (*Current Protocols in Molecular Biology*, Ausubel et al., eds., John Wiley & Sons, 1992). Other examples of eukaryotic heterologous signal sequences include the secretory sequences of

melittin and human placental alkaline phosphatase (Stratagene, La Jolla, California). In yet another example, useful prokaryotic heterologous signal sequences include the phoA secretory signal (*Molecular cloning*, Sambrook et al, second edition, Cold spring harbor laboratory press, 1989) and the protein A secretory signal (Pharmacia Biotech, Piscataway, New Jersey)

In yet another embodiment, the fusion protein is an BGCKr-immunoglobulin fusion protein in which all or part of BGCKr is fused to sequences derived from a member of the immunoglobulin protein family. The BGCKr-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a BGCKr ligand and a BGCKr protein on the surface of a cell, to thereby suppress BGCKr-mediated signal transduction *in vivo*. The BGCKr-immunoglobulin fusion proteins can be used to affect the bioavailability of a BGCKr cognate ligand. Inhibition of the BGCKr ligand/BGCKr interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (e.g. promoting or inhibiting) cell survival. Moreover, the BGCKr-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-BGCKr antibodies in a subject, to purify BGCKr ligands and in screening assays to identify molecules which inhibit the interaction of BGCKr with a BGCKr ligand.

Preferably, a BGCKr chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for

appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g., Current Protocols in Molecular Biology*, Ausubel et al. eds., John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). An BGCKr-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the BGCKr protein.

The present invention also pertains to variants of the BGCKr proteins which function as either BGCKr agonists (mimetics) or as BGCKr antagonists. Variants of the BGCKr protein can be generated by mutagenesis, e.g., discrete point mutation or truncation of the BGCKr protein. An agonist of the BGCKr protein can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of the BGCKr protein. An antagonist of the BGCKr protein can inhibit one or more of the activities of the naturally occurring form of the BGCKr protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the BGCKr protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein can have fewer side effects in a subject

relative to treatment with the naturally occurring form of the BGCKr proteins.

Variants of the BGCKr protein which function as either BGCKr agonists (mimetics) or as BGCKr antagonists can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the BGCKr protein for BGCKr protein agonist or antagonist activity. In one embodiment, a variegated library of BGCKr variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of BGCKr variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential BGCKr sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of BGCKr sequences therein. There are a variety of methods which can be used to produce libraries of potential BGCKr variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential BGCKr sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (*see, e.g.*, Narang (1983) *Tetrahedron* 39:3; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477).

In addition, libraries of fragments of the BGCKr protein coding sequence can be used to generate a variegated population of BGCKr fragments for screening and subsequent selection of variants of a BGCKr protein.

In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a BGCKr coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the BGCKr protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of BGCKr proteins. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recrusive ensemble mutagenesis (REM), a technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify BGCKr variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave et al. (1993) *Protein Engineering* 6(3):327-331).

In one embodiment, cell based assays can be exploited to analyze a variegated BGCKr library. For example, a library of expression vectors can be transfected into a cell line which ordinarily responds to a particular ligand or activating factor in an BGCKr-dependent manner. The transfected cells are then contacted with the chemokine receptor and the effect of expression of the mutant on signaling by the chemokine receptor can be detected. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of cytokine induction, and the individual clones further characterized.

An isolated BGCKr protein, or a portion or fragment. thereof, can be used as an immunogen to generate antibodies that bind BGCKr using standard techniques for polyclonal and monoclonal antibody preparation. The full-length BGCKr protein can be used or, alternatively, the invention provides antigenic peptide fragments of BGCKr for use as immunogens. The antigenic peptide of BGCKr comprises at least 8 (preferably 10, 15, 20, or 30) amino acid residues of the amino acid sequence shown in SEQ ID NO:5 and encompasses an epitope of BGCKr such that an antibody raised against the peptide forms a specific immune complex with BGCKr.

Preferred epitopes encompassed by the antigenic peptide are regions of BGCKr that are located on the surface of the protein, e.g., hydrophilic regions. A hydrophobicity analysis of the human BGCKr protein sequence indicates that the regions between amino acids 22-36 and 180-197 are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production.

A BGCKr immunogen typically is used to prepare antibodies by immunizing a suitable subject, (e.g., rabbit, goat, mouse or other mammal) with the immunogen.

An appropriate immunogenic preparation can contain, for example, recombinantly expressed BGCKr protein or a chemically synthesized BGCKr polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic BGCKr preparation induces a polyclonal anti-BGCKr antibody response.

Accordingly, another aspect of the invention pertains to anti-BGCKr antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as BGCKr. A molecule which specifically binds to BGCKr is a molecule which binds BGCKr, but does not substantially bind other molecules in a sample, e.g., a biological sample, which naturally contains BGCKr. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind BGCKr. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of BGCKr. A monoclonal antibody composition thus typically displays a single binding affinity for a particular BGCKr protein with which it immunoreacts.

Polyclonal anti-BGCKr antibodies can be prepared as described above by immunizing a suitable subject with a BGCKr immunogen. The anti-BGCKr antibody titer in the immunized subject can be monitored over time by standard

techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized BGCKr. If desired, the antibody molecules directed against BGCKr can be isolated from the mammal (e.g., from the blood) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, e.g., when the anti-BGCKr antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497, the human B cell hybridoma technique (Kozbor et al. (1983) *Immunol Today* 4:72), the EBV-hybridoma technique (Cole et al. (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing various antibodies monoclonal antibody hybridomas is well known (*see generally* Current Protocols in Immunology (1994) Coligan et al. (eds.) John Wiley & Sons, Inc., New York, NY). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a BGCKr immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds BGCKr.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-BGCKr monoclonal antibody (see Current Protocols in Immunology, *supra*) (*see, e.g.*, Galfre et al. (1977) *Nature* 266:55052. Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (e.g., a myeloma cell line) is derived from the same

mammalian species as the lymphocytes. For example, murine hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line, e.g., a myeloma cell line that is sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, e.g., the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind BGCKr, e.g., using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-BGCKr antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with BGCKr to thereby isolate immunoglobulin library members that bind BGCKr. Kits for generating and screening phage display libraries are commercially available (e.g., the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, Ladner et al. U.S. Patent No. 5,223,409; PCT Publication

No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-1372; Hay et al. (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse et al. (1989) *Science* 246:1275-1281; Griffiths et al. (1993) *EMBO J* 12:725-734.

Additionally, recombinant anti-BGCKr antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in PCT Publication No. WO 87/02671; European Patent Application 184,187; European Patent Application 171,496; European Patent Application 173,494; PCT Publication No. WO 86/01533; U.S. Patent No. 4,816,567; European Patent Application 125,023; Better et al. (1988) *Science* 240:1041-1043; Liu et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu et al. (1987) *J. Immunol.* 139:3521-3526; Sun et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura et al. (1987) *Canc. Res.* 47:999-1005; Wood et al. (1985) *Nature* 314:446-449; and Shaw et al. (1988) *J. Natl. Cancer Inst.* 80:1553-1559); Morrison, (1985) *Science* 229:1202-1207; Oi et al. (1986) *Bio/Techniques* 4:214; U.S. Patent 5,225,539; Jones et al. (1986) *Nature* 321:552-525; Verhoeyan et al. (1988) *Science* 239:1534; and Beidler et al. (1988) *J. Immunol.* 141:4053-4060.

An anti-BGCKr antibody (e.g., monoclonal antibody) can be used to isolate BGCKr by standard techniques, such as affinity chromatography or immunoprecipitation. An

anti-BGCKr antibody can facilitate the purification of natural BGCKr from cells and of recombinantly produced BGCKr expressed in host cells. Moreover, an anti-BGCKr antibody can be used to detect BGCKr protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the BGCKr protein. Anti-BGCKr antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

III. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding BGCKr (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a

"plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors, expression vectors, are capable of directing the expression of genes to which they are operatively linked. In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids (vectors). However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The

term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., BGCKr proteins, mutant forms of BGCKr, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of BGCKr in prokaryotic or eukaryotic cells, e.g., bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins.

Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., (1988) *Gene* 69:301-315) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada et al. (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the BGCKr expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1 (Baldari et al. (1987) *EMBO J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz et al. (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and picZ (InVitrogen Corp, San Diego, CA).

Alternatively, BGCKr can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al. (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed (1987) *Nature* 329:840) and pMT2PC (Kaufman et al. (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral

regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook et al. (*supra*).

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al. (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA

molecule) of an RNA molecule which is antisense to BGCKr mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes See Weintraub et al., *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, BGCKr protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*supra*), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding BGCKr or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) BGCKr protein. Accordingly, the invention further provides methods for producing BGCKr protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a

recombinant expression vector encoding BGCKr has been introduced) in a suitable medium such that BGCKr protein is produced. In another embodiment, the method further comprises isolating BGCKr from the medium or the host cell.

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which BGCKr-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous BGCKr sequences have been introduced into their genome or identical recombinant animals in which endogenous BGCKr sequences have been altered. Such animals are useful for studying the function and/or activity of BGCKr and for identifying and/or evaluating modulators of BGCKr activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, an "identical recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous BGCKr gene has been altered by identical recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing BGCKr-encoding nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The BGCKr cDNA sequence e.g., that of (SEQ ID NO:1, 3, 4, 6, 7, or the cDNA of ATCC 209756) can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of the human BGCKr gene, such as a mouse BGCKr gene, can be isolated based on hybridization to the human BGCKr cDNA and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the BGCKr transgene to direct expression of BGCKr protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, U.S. Patent No. 4,873,191 and in Hogan, *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the BGCKr transgene in its genome and/or expression of BGCKr mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding BGCKr can further be bred to other transgenic animals carrying other transgenes.

To create an identical recombinant animal, a vector is prepared which contains at least a portion of a

BGCKr gene (e.g., a human or a non-human homologous of the BGCKr gene, e.g., murine BGCKr gene) into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the BGCKr gene. In a preferred embodiment, the vector is designed such that, upon identical recombination, the endogenous BGCKr gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon identical recombination, the endogenous BGCKr gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous BGCKr protein). In the identical recombination vector, the altered portion of the BGCKr gene is flanked at its 5' and 3' ends by additional nucleic acid of the BGCKr gene to allow for identical recombination to occur between the exogenous BGCKr gene carried by the vector and an endogenous BGCKr gene in an embryonic stem cell. The additional flanking BGCKr nucleic acid is of sufficient length for successful identical recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (*see e.g.*, Thomas and Capecchi (1987) *Cell* 51:503 for a description of identical recombination vectors). The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced BGCKr gene has identically recombined with the endogenous BGCKr gene are selected (*see e.g.*, Li et al. (1992) *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (*see, e.g.*, Bradley in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, ed. (IRL, Oxford, 1987) pp. 113-

152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the identically recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the identically recombined DNA by germline transmission of the transgene. Methods for constructing identical recombination vectors and identical recombinant animals are described further in Bradley (1991) *Current Opinion in Bio/technology* 2:823-829 and in PCT Publication Nos. WO 90/11354, WO 91/01140, WO 92/0968, and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, *see, e.g.*, Lakso et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman et al. (1991) *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut et al. (1997) *Nature* 385:810-813 and PCT Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the

transgenic animal can be isolated and induced to exit the growth cycle and enter G_o phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

IV. Pharmaceutical Compositions

The BGCKr nucleic acid molecules, BGCKr proteins, and anti-BGCKr antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal,

subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF; Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper

fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a BGCKr protein or anti-BGCKr antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral

compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring. For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases

such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see* U.S. Patent 5,328,470) or by stereotactic injection (*see e.g.*, Chen et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

V. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) detection assays (e.g., chromosomal mapping, tissue typing, forensic biology), c) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenomics); and d) methods of treatment (e.g., therapeutic and prophylactic). In various embodiments the BGCKr protein can have one or more of the following activities: (a) the ability to interact on the cell suface with a second non-BGCKr molecule on the surface of the same cell, (b) the ability to interact on the cell surface with a second non-BGCKr protein molecule on the surface of a different cell, (c) the ability to form a complex with a ligand e.g., a

chemokine or peptide, (d) the ability to interact with an intracellular protein including protein kinase or cytoskeletal protein, and (e) the ability to stimulate cellular migration. BGCKr nucleic acid molecules, proteins, protein homologues, and antibodies may also used to modulate cellular signal transduction either *in vivo* or *in vitro*, modulate pro-inflammatory functions of chemokines, modulate stimulatory functions of chemokines, regulate cellular proliferation, regulate cellular migration, regulate cell adhesion, regulate cellular homing (e.g., during an inflamatory response) and regulate exocytosis.

The isolated nucleic acid molecules of the invention can be used to express BGCKr protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect BGCKr mRNA (e.g., in a biological sample) or a genetic lesion in a BGCKr gene, and to modulate BGCKr activity. In addition, the BGCKr proteins can be used to screen drugs or compounds which modulate the BGCKr activity or expression as well as to treat disorders characterized by insufficient or excessive production of BGCKr protein or production of BGCKr protein forms which have decreased or aberrant activity compared to BGCKr wild type protein. In addition, the anti-BGCKr antibodies of the invention can be used to detect and isolate BGCKr proteins and modulate BGCKr activity.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

A. Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents

(e.g., peptides, peptidomimetics, small molecules or other drugs) which bind to BGCKr proteins or have a stimulatory or inhibitory effect on, for example, BGCKr expression or BGCKr activity.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of a BGCKr protein or polypeptide or biologically active portion thereof. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al. (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann et al. (1994). *J. Med. Chem.* 37:2678; Cho et al. (1993) *Science* 261:1303; Carrell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and Gallop et al. (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Bio/Techniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner U.S. Patent No. 5,223,409), spores (Ladner U.S. Patent No.

5,571,698; 5,403,484 and 5,223,409), plasmids (Cull et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390; Devlin (1990) *Science* 249:404-406; Cwirla et al. (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382; and Felici (1991) *J. Mol. Biol.* 222:301-310).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of BGCKr protein, or a biologically active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to a BGCKr protein determined. The cell, for example, can be a yeast cell or a cell of mammalian origin. Determining the ability of the test compound to bind to the BGCKr protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the BGCKr protein or biologically active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In a preferred embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of BGCKr protein, or a biologically active portion thereof, on the cell surface with a known compound which binds BGCKr to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a BGCKr protein, wherein determining the ability of

the test compound to interact with a BGCKr protein comprises determining the ability of the test compound to preferentially bind to BGCKr or a biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of BGCKr protein, or a biologically active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the BGCKr protein or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of BGCKr or a biologically active portion thereof can be accomplished, for example, by determining the ability of the BGCKr protein to bind to or interact with a BGCKr target molecule. As used herein, a "target molecule" is a molecule with which a BGCKr protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a BGCKr protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A BGCKr target molecule can be a non-BGCKr molecule or a BGCKr protein or polypeptide of the present invention. In one embodiment, a BGCKr target molecule is a component of a signal transduction pathway which facilitates transduction of an extracellular signal (e.g., a signal generated by binding of a compound to a membrane-bound BGCKr molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein which has catalytic activity or a protein which facilitates the association of downstream signaling molecules with BGCKr.

Determining the ability of the BGCKr protein to bind to or interact with a BGCKr target molecule can be

accomplished by one of the methods described above for determining direct binding. In a preferred embodiment, determining the ability of the BGCKr protein to bind to or interact with a BGCKr target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (e.g., intracellular Ca^{2+}, diacylglycerol, IP3, etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (e.g., a BGCKr-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, e.g. luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the present invention is a cell-free assay comprising contacting a BGCKr protein or biologically active portion thereof with a test compound and determining the ability of the test compound to bind to the BGCKr protein or biologically active portion thereof. Binding of the test compound to the BGCKr protein can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the BGCKr protein or biologically active portion thereof with a known compound which binds BGCKr to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a BGCKr protein, wherein determining the ability of the test compound to interact with a BGCKr protein comprises determining the ability of the test compound to preferentially bind to BGCKr or biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-free assay comprising contacting BGCKr protein or biologically active portion thereof with a test compound and determining the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the BGCKr protein or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of BGCKr can be accomplished, for example, by determining the ability of the BGCKr protein to bind to a BGCKr target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of BGCKr can be accomplished by determining the ability of the BGCKr protein further modulate a BGCKr target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as previously described.

In yet another embodiment, the cell-free assay comprises contacting the BGCKr protein or biologically active portion thereof with a known compound which binds BGCKr to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a BGCKr protein, wherein determining the ability of the test compound to interact with a BGCKr protein comprises determining the ability of the BGCKr protein to preferentially bind to or modulate the activity of a BGCKr target molecule.

The cell-free assays of the present invention are amenable to use of both the soluble form or the membrane-bound form of BGCKr. In the case of cell-free assays comprising the membrane-bound form of BGCKr, it may be desirable to utilize a solubilizing agents such that the membrane-bound form of BGCKr is maintained in solution. Examples of such solubilizing agents include non-ionic

detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either BGCKr or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to BGCKr, or interaction of BGCKr with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/ BGCKr fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical; St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or BGCKr protein, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the

complexes can be dissociated from the matrix, and the level of BGCKr binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either BGCKr or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated BGCKr or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals; Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with BGCKr or target molecules but which do not interfere with binding of the BGCKr protein to its target molecule can be derivatized to the wells of the plate, and unbound target or BGCKr trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the BGCKr or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the BGCKr or target molecule.

In another embodiment, modulators of BGCKr expression are identified in a method in which a cell is contacted with a candidate compound and the expression of BGCKr mRNA or protein in the cell is determined. The level of expression of BGCKr mRNA or protein in the presence of the candidate compound is compared to the level of expression of BGCKr mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of BGCKr expression based on this comparison. For example, when expression of BGCKr mRNA or protein is greater

(statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of BGCKr mRNA or protein expression. Alternatively, when expression of BGCKr mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of BGCKr mRNA or protein expression. The level of BGCKr mRNA or protein expression in the cells can be determined by methods described herein for detecting BGCKr mRNA or protein.

In yet another aspect of the invention, the BGCKr proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al. (1993) *Bio/Techniques* 14:920-924; Iwabuchi et al. (1993) *Oncogene* 8:1693-1696; and WO94/10300), to identify other proteins, which bind to or interact with BGCKr ("BGCKr-binding proteins" or "BGCKr-bp") and modulate BGCKr activity. Such BGCKr-binding proteins are also likely to be involved in the propagation of signals by the BGCKr proteins as, for example, upstream or downstream elements of the BGCKr pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for BGCKr is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If

the "bait" and the "prey" proteins are able to interact, *in vivo*, forming an BGCKr-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with BGCKr.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

B. Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

1. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. Accordingly, BGCKr nucleic acid molecules described herein or fragments thereof, can be used to map the location of BGCKr genes on a chromosome. The mapping of the BGCKr sequences to chromosomes is an important first

step in correlating these sequences with genes associated with disease.

Briefly, BGCKr genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the BGCKr sequences. Computer analysis of BGCKr sequences can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the BGCKr sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (e.g., human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but human cells can, the one human chromosome that contains the gene encoding the needed enzyme, will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. (D'Eustachio et al. (1983) *Science* 220:919-924). Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the BGCKr sequences to design oligonucleotide primers,

sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a BGCKr sequence to its chromosome include *in situ* hybridization (described in Fan et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma et al., Human Chromosomes: A Manual of Basic Techniques (Pergamon Press, New York, 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene

families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, e.g., Egeland et al. (1987) *Nature*, 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the BGCKr gene can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

2. Tissue Typing

The BGCKr sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its

personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the BGCKr sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The BGCKr sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate

individuals. The noncoding sequences of SEQ ID NO:1 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:3 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from BGCKr sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

3. Use of Partial BGCKr Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that

is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:4 are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the BGCKr sequences or portions thereof, e.g., fragments derived from the noncoding regions of SEQ ID NO:4 having a length of at least 20 or 30 bases.

The BGCKr sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labeable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, e.g., brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such BGCKr probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, e.g., BGCKr primers or probes can be used to screen tissue culture for contamination (i.e., screen for the presence of a mixture of different types of cells in a culture).

C. Predictive Medicine:

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trails are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining BGCKr protein and/or nucleic acid expression as well as BGCKr activity, in the

context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant BGCKr expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with BGCKr protein, nucleic acid expression or activity. For example, mutations in a BGCKr gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with BGCKr protein, nucleic acid expression or activity.

Another aspect of the invention provides methods for determining BGCKr protein, nucleic acid expression or BGCKr activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (e.g., drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (e.g., the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (e.g., drugs or other compounds) on the expression or activity of BGCKr in clinical trials.

These and other agents are described in further detail in the following sections.

1. Diagnostic Assays

An exemplary method for detecting the presence or absence of BGCKr in a biological sample involves

obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting BGCKr protein or nucleic acid (e.g., mRNA, genomic DNA) that encodes BGCKr protein such that the presence of BGCKr is detected in the biological sample. A preferred agent for detecting BGCKr mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to BGCKr mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length BGCKr nucleic acid, such as the nucleic acid of SEQ ID NO: 1, 3, 4, 6, 7, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to BGCKr mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

A preferred agent for detecting BGCKr protein is an antibody capable of binding to BGCKr protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a

subject. That is, the detection method of the invention can be used to detect BGCKr mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of BGCKr mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of BGCKr protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of BGCKr genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of BGCKr protein include introducing into a subject a labeled anti-BGCKr antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting BGCKr protein, mRNA, or genomic DNA, such that the presence of BGCKr protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of BGCKr protein, mRNA or genomic DNA in the control sample with the presence of BGCKr protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of BGCKr in a biological sample (a test sample). Such kits can be used to determine if a subject

is suffering from or is at increased risk of developing a disorder associated with aberrant expression of BGCKr (e.g., an immunological disorder). For example, the kit can comprise a labeled compound or agent capable of detecting BGCKr protein or mRNA in a biological sample and means for determining the amount of BGCKr in the sample (e.g., an anti-BGCKr antibody or an oligonucleotide probe which binds to DNA encoding BGCKr, e.g., SEQ ID NO:1, 3, 4, 6, or 7). Kits may also include instruction for observing that the tested subject is suffering from or is at risk of developing a disorder associated with aberrant expression of BGCKr if the amount of BGCKr protein or mRNA is above or below a normal level.

For antibody-based kits, the kit may comprise, for example: (1) a first antibody (e.g., attached to a solid support) which binds to BGCKr protein; and, optionally, (2) a second, different antibody which binds to BGCKr protein or the first antibody and is conjugated to a detectable agent.

For oligonucleotide-based kits, the kit may comprise, for example: (1) a oligonucleotide, e.g., a detectably labelled oligonucleotide, which hybridizes to a BGCKr nucleic acid sequence or (2) a pair of primers useful for amplifying a BGCKr nucleic acid molecule;

The kit may also comprise, e.g., a buffering agent, a preservative, or a protein stabilizing agent. The kit may also comprise components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit may also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit is usually enclosed within an individual container and all of the various containers are within a single package along with instructions for observing

whether the tested subject is suffering from or is at risk of developing a disorder associated with aberrant expression of BGCKr.

2. Prognostic Assays

The methods described herein can furthermore be utilized as diagnostic or prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with aberrant BGCKr expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with BGCKr protein, nucleic acid expression or activity such as autoimmune diseases. Thus, the present invention provides a method in which a test sample is obtained from a subject and BGCKr protein or nucleic acid (e.g., mRNA, genomic DNA) is detected, wherein the presence of BGCKr protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant BGCKr expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (e.g., serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant BGCKr expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with a specific agent or class of agents (e.g., agents of a type which decrease BGCKr activity). Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant BGCKr expression or activity in which a test sample is obtained and BGCKr

protein or nucleic acid is detected (e.g., wherein the presence of BGCKr protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant BGCKr expression or activity).

The methods of the invention can also be used to detect genetic lesions or mutations in a BGCKr gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding a BGCKr-protein, or the mis-expression of the BGCKr gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from a BGCKr gene; 2) an addition of one or more nucleotides to a BGCKr gene; 3) a substitution of one or more nucleotides of a BGCKr gene, 4) a chromosomal rearrangement of a BGCKr gene; 5) an alteration in the level of a messenger RNA transcript of a BGCKr gene, 6) aberrant modification of a BGCKr gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a BGCKr gene, 8) a non-wild type level of a BGCKr-protein, 9) allelic loss of a BGCKr gene, and 10) inappropriate post-translational modification of a BGCKr-protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a BGCKr gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran et al. (1988) *Science* 241:1077-1080; and Nakazawa et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the BGCKr-gene (see Abravaya et al. (1995) *Nucleic Acids Res.* 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a BGCKr gene under conditions such that hybridization and amplification of the BGCKr-gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi et al. (1988) *Bio/Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of

nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a BGCKr gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.*, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in BGCKr can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin et al. (1996) *Human Mutation* 7:244-255; Kozal et al. (1996) *Nature Medicine* 2:753-759). For example, genetic mutations in BGCKr can be identified in two-dimensional arrays containing light-generated DNA probes as described in Cronin et al. supra. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to

the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the BGCKr gene and detect mutations by comparing the sequence of the sample BGCKr with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert ((1977) *Proc. Natl. Acad. Sci. USA* 74:560) or Sanger ((1977) *Proc. Natl. Acad. Sci. USA* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Bio/Techniques* 19:448), including sequencing by mass spectrometry (*see, e.g.*, PCT Publication No. WO 94/16101; Cohen et al. (1996) *Adv. Chromatogr.* 36:127-162; and Griffin et al. (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in the BGCKr gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers et al. (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type BGCKr sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and

with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, e.g., Cotton et al (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba et al (1992) *Methods Enzymol.* 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in BGCKr cDNAs obtained from samples of cells. For example, the mutY enzyme of E. coli cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu et al. (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on a BGCKr sequence, e.g., a wild-type BGCKr sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.*, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in BGCKr genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita et al. (1989) *Proc Natl. Acad. Sci USA*: 86:2766, *see also* Cotton (1993) *Mutat. Res.* 285:125-144; and Hayashi (1992) *Genet Anal Tech Appl* 9:73-79). Single-stranded DNA fragments of sample and control BGCKr nucleic acids will be denatured and allowed to renature. The secondary structure of

single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen et al. (1991) *Trends Genet* 7:5).

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers et al. (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki et al. (1986) *Nature* 324:163); Saiki et al. (1989) *Proc. Natl Acad. Sci USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of

different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs et al. (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition, it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini et al. (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, e.g., in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a BGCKr gene.

Furthermore, any cell type or tissue, preferrably peripheral blood leukocytes, in which BGCKr is expressed

may be utilized in the prognostic assays described herein.

3. Pharmacogenomics

Agents, or modulators which have a stimulatory or inhibitory effect on BGCKr activity (e.g., BGCKr gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (e.g., immunological disorders) associated with aberrant BGCKr activity. In conjunction with such treatment, the pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (e.g., drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of BGCKr protein, expression of BGCKr nucleic acid, or mutation content of BGCKr genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Linder (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions

transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (e.g., N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular

basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of BGCKr protein, expression of BGCKr nucleic acid, or mutation content of BGCKr genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a BGCKr modulator, such as a modulator identified by one of the exemplary screening assays described herein.

4. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of BGCKr (e.g., the ability to modulate aberrant cell proliferation and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase BGCKr gene expression, protein levels, or upregulate BGCKr activity, can be monitored in clinical trails of subjects exhibiting decreased BGCKr gene expression, protein levels, or downregulated BGCKr activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease BGCKr gene expression, protein levels, or downregulated BGCKr activity, can be monitored in clinical trails of subjects exhibiting increased BGCKr gene expression, protein levels, or upregulated BGCKr activity. In such clinical trials, the expression or

activity of BGCKr and, preferably, other genes that have been implicated in, for example, a cellular proliferation disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

For example, and not by way of limitation, genes, including BGCKr, that are modulated in cells by treatment with an agent (e.g., compound, drug or small molecule) which modulates BGCKr activity (e.g., identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of BGCKr and other genes implicated in the disorder. The levels of gene expression (i.e., a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of BGCKr or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a BGCKr protein, mRNA, or genomic DNA in the preadministration sample;

(iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the BGCKr protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the BGCKr protein, mRNA, or genomic DNA in the pre-administration sample with the BGCKr protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of BGCKr to higher levels than detected, i.e., to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of BGCKr to lower levels than detected, i.e., to decrease the effectiveness of the agent.

C. Methods of Treatment:

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant BGCKr expression or activity.

1. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant BGCKr expression or activity, by administering to the subject an agent which modulates BGCKr expression or at least one BGCKr activity. Subjects at risk for a disease which is caused or contributed to by aberrant BGCKr expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein.

Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the BGCKr aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of BGCKr aberrancy, for example, a BGCKr agonist or BGCKr antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

2. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating BGCKr expression or activity for therapeutic purposes. For example, since chemokines play a crucial role in recruitment and activation of leukocytes, one can, by administering a therapeutically effective dose ameliorate symptoms of disorders associated with aberrant inflammation. Moreover, modulating BGCKr expression or activity may play a role in treating angiogenesis, tumor growth, proliferation, allergic reactions and the entry of human immunodeficiency virus (HIV) into cells. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of BGCKr protein activity associated with the cell. An agent that modulates BGCKr protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a BGCKr protein, a peptide, a BGCKr peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more of the biological activities of BGCKr protein. Examples of such stimulatory agents include active BGCKr protein and a nucleic acid molecule encoding BGCKr that has been introduced into the cell. In another embodiment, the agent inhibits one or more of the biological activities of BGCKr protein. Examples of such

inhibitory agents include antisense BGCKr nucleic acid molecules and anti-BGCKr antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g, by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a BGCKr protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) BGCKr expression or activity. In another embodiment, the method involves administering a BGCKr protein or nucleic acid molecule as therapy to compensate for reduced or aberrant BGCKr expression or activity.

Stimulation of BGCKr activity is desirable in situations in which BGCKr is abnormally downregulated and/or in which increased BGCKr activity is likely to have a beneficial effect. Conversely, inhibition of BGCKr activity is desirable in situations in which BGCKr is abnormally upregulated and/or in which decreased BGCKr activity is likely to have a beneficial effect.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are hereby incorporated by reference.

EXAMPLES

Example 1: Isolation and Characterization of Human BGCKr cDNAs

The human amino acid sequences for human CC chemokine receptors CC-1, CC-2, CC-3, CC-4, CC-5, CC-6 and CC-7 were used to search proprietary EST and dbEST databases using TBLASTN (Washington University; version 2.0, BLOSUM62 search matrix), and all sequences scoring better than 60 were saved. Sequences identified in this manner were assembled using Phrap (P.Green, University of Washington) and the sequences from the assembly compared back to the public and proprietary protein databases using BLASTX (Washington University version 2.0, BLOSUM62 search matrix). Sequences exhibiting 90% or better identity to any protein present in Genpept, SwissProt, or PIR were marked as examples of these proteins and removed. The remaining sequences were compared back to the proprietary EST and dbEST databases using BLASTN (Washington University; version 2.0, mismatch score of -10, minimum score of 300) to identify overlapping ESTs.

This search led to the identification of an EST (Genbank Accession Number N67224; originally found in a human Weizmann Olfactory Epithelium library) sequence found in dbEST which was thought likely to encode a portion of a gene having some similarity to previously identified chemokine receptors.

A cDNA encoding the BGCKr was identified using the following procedure. First, a BGCKr DNA probe was constructed using specific forward (F) and reverse (R) primers that were designed based on the EST sequence.

F: 5'GATTTACTCCTTCTATTCACTCTG 3' (SEQ ID NO:17)

R: 5'GCACCTAGCATTGTCATTTACTG 3' (SEQ ID NO:18)

Amplification was done in a thermocycler (MJ Research; Watertown, MA) using Taq polymerase (Perkin

Elmer) with a 51°C annealing temperature (92°C for 30 sec, 51°C for 30 sec, and 42°C for 45 sec). The template was 200 ng of human genomic DNA and it gave rise to the expected 294 bp band. Since the PCR product sequenced accurately, it was used as a probe.

Ten human cDNA libraries were tested by PCR with the primers and conditions described above. Three of the libraries (human lymph node, human congestive heart failure (CHF), and human thymus) showed the same 294 bp band. Sequencing analysis revealed that the same cDNA fragment was amplified from these libraries.

The human CHF library was screened to identify a full length BGCKr cDNA using Boehringer Mannheim's DIG labeling kit, (Catalog Number 1093 657), nylon membranes (Catalog Number 1699 083), and hybridization solution (Catalog Number 1603 558), according to the manufacturer's instructions. This kit uses a non-radioactive, enzymatically digoxigenin-labeled probe, that is detected using a chemiluminescent alkaline phosphatase substrate. 5 X 10^5 phage were plated per 140 mm plate in each of ten plates. Four positive plaques were found in the primary screen. After an overnight elution into SM buffer (100 mM NaCl; 8 mM MgSO4; 5 mM Tris, pH 7.5; 0.01% gelatin), the eluates were diluted to 10^{-3} and plated 1:10 onto each of 4 new plates. Five positive clones were found in this second round of screening. Following elution of those five phage plaques, the plasmid was excised with DH10B (ZIP) cells (GIBCO/BRL; Catalog Number Y02043) and plated onto IPTG/X-Gal/LB/Amp plates. Twelve clones from those plates were sequenced.

Four clones out of the twelve were then fully sequenced. Three of them were identical, and one was shorter. The BGCKr cDNA (SEQ ID NO:1) has a 999 nucleotide open reading frame (nucleotides 19-1017 of SEQ

ID NO:1; SEQ ID NO:3) which encodes a 333 amino acid protein (SEQ ID NO:2),

The cDNA shown in Figure 1 (SEQ ID NO:1) proved to be a partial cDNA clone.

An alignment of the obtained sequence with the published bovine gustatory receptor Type B (Accession Number S63848), suggested the presence of additional 5' coding sequence. The gene was therefore RACEd using Marathon-Ready fetal spleen cDNA (Clontech Laboratories, Inc., Palo Alto, CA). The gene-specific primers used were:

GSP1: 5' cttttgattctgagggtcctaca 3' (SEQ ID NO:22)

GSP2: 5' cagtaaatgacaatgctaggtgc 3' (SEQ ID NO:23)

PCR was performed per manufacturer's protocol with Clontech's Advantage cDNA polymerase mix. Nested PCR was performed with GSP2 using the same PCR conditions.

The resulting product was cloned into pGEM-T (Promega Corp., Madison, WI) and 12 clones were sequenced. A novel 5' sequence was identified.

Two primers were designed and used to amplify the entire gene from human lymph node marathon-ready cDNA library (Stratagene).

Forward primer: 5' CCCCCCGAATTCTCTCTGCCGACTACAACA 3' (SEQ ID NO:9)

Reverse primer: 5' CCCCCCAAGCTTATCCAAGCAAAAGGCAGAGCA 3' (SEQ ID NO:24)

The forward primer contained an Eco R1 site, and the reverse primer contained a Hind 3 site, which were used for cloning into pBluescript (Stratagene, La Jolla, CA).

A cDNA (SEQ ID NO:4) encoding full-length human BGCKr is shown in Figure 2. The 1050 nucleotide open reading frame (SEQ ID NO:6) of this cDNA encodes a 350 amino acid protein (SEQ ID NO:5)

To obtain the full length sequence of the murine homologue of human BGCKr, m2038, a probe based on Accession Numbers AA050273 and AA014373, was used to screen a BAC library of mouse genomic DNA (RPCI-22 strain 129 mouse BAC library from BAC-PAC Resources, Department of Cancer Genetics, Roswell Park Cancer Institute, Buffalo, NY). The open reading frame and translation product of the sequence obtained is shown in Figure 3. This cDNA (SEQ ID NO:7) encodes a 350 amino acid protein (SEQ ID NO:8).

Example 2: Characterization of BGCKr Proteins

In this example, the predicted amino acid sequence of human BGCKr protein was compared to amino acid sequences of known proteins and various motifs were identified. In addition, the molecular weight of the human BGCKr protein was predicted.

The human BGCKr cDNA isolated as described above (Figure 2; SEQ ID NO:4) encodes a 350 amino acid protein (Figure 1; SEQ ID NO:2). Human BGCKr has a predicted MW of 39.9 kDa, not including post-translational modifications. It has a predicted isoelectric point of 8.518 and a predicted 7.878 charge at pH 7.0.

Within the BGCKr amino acid sequence (SEQ ID NO:5) potential N-glycosylation sites are present at amino acids 6-9, 19-22, and 276-279; potential protein kinase C phosphorylation sites are present at amino acids 195-197, 223-225, 278-280, 309-311, and 323-325; potential casein kinase II phosphorylation sites are present at amino acids 25-28, 74-77, 177-180, and 330-333; a potential tycosine kinase phosphorylation site is present at amino acids 263-269, a potential N-myristoylation site is present at amino acids 55-60, and a G-protein coupled receptor signature sequence is present at amino acids 125-141.

Figure 5 depicts an alignment of the amino acid sequences of full-length human BGCKr, bovine gustatory receptor (SEQ ID NO:10, accesssion number P35350), ceraeBonzo (SEQ ID NO:11, accession number 018983), human CCR6 (SEQ ID NO:12, accession number P51684), human CCR7 (SEQ ID NO:13, accession number P32248), human GPR-9-6 (SEQ ID NO:14, accession number P51686), murine CCR6 (SEQ ID NO:15, accession number AJ222714), and murine CCR7 (SEQ ID NO:16, accession number W48723).

Figure 6 depicts an alignment of the amino acid sequence of murine BGCKr with bovine gustatory receptor (SEQ ID NO:10, accession number P35350) and human BGCKr (SEQ ID NO:5).

Figure 10 depicts an alignment of a portion of human BGCKr (amino acids 58-303) with a seven transmembrane receptor consensus sequence.

Example 3: Preparation of BGCKr Proteins

Recombinant BGCKr can be produced in a variety of expression systems. For example, the mature BGCKr peptide can be expressed as a recombinant glutathione-S-transferase (GST) fusion protein in *E. coli* and the fusion protein can be isolated and characterized. Specifically, as described above, BGCKr can be fused to GST and this fusion protein can be expressed in *E. coli* strain PEB199. As BGCKr is predicted to be 110 kD and GST is predicted to be 26 kD, the fusion protein is predicted to be 136 kD in molecular weight. Expression of the GST-BGCKr fusion protein in PEB199 can be induced with IPTG. The recombinant fusion protein can be purified from crude bacterial lysates of the induced PEB199 strain by affinity chromatography on glutathione beads. Using polyacrylamide gel electrophoretic analysis of the proteins purified from the bacterial lysates, the resultant fusion protein should be 136 kD in size.

Example 4: Anti-BGCKr antibodies

Antibodies to the second and third extracellular loops as well as to the amino terminal domain of this receptor can be prepared. These antibodies can be used for *in vivo* and *in vitro* blockage experiments and protein expression analysis by immunofluorescence assay (FACS) and/or immunohistological staining.

Preferred peptides for antibody generation include:

YDYSQYELICIKEDVREFAKV (SEQ ID NO:25; amino acid 5 to amino acid 25 of SEQ ID NO:2) and IKEDVREFAKV (SEQ ID NO:26; amino acid 15 to amino acid 25 of SEQ ID NO:2) (amino terminal peptides); DNARCIPIFPRYLGTSMK (SEQ ID NO:27; amino acid 163 to amino acid 180 of SEQ ID NO:2) (second extracellular loop sequence); and MSKRMDI (SEQ ID NO:28; amino acid 160 to amino acid 166 of SEQ ID NO:2) (third extracellular loop sequence).

Antibodies were generated using the following peptides: IKEDVR (SEQ ID NO:26); DNARCI (SEQ ID NO:29) and MSKRMDIA (SEQ ID NO:30). These peptides are from the N-terminus, second extracellular loop, and third extracellular loop, respectively of human BGCKr (SEQ ID NO:5).

Example 5: BCGKr a potential co-receptor for human immunodeficiency virus (HIV).

The entry of HIV into a cell requires both CD4 and a G-protein coupled receptor. A variety of approaches can be used to understand whether BGCKr can act as a co-receptor for HIV.

For example binding studies can be conducted to determine whether HIV envelope proteins interact with BGCKr. An antibody directed against either a HIV envelope protein (for example, gp120), or BGCKr can be used in an immunoprecipitation study to determine whether

BGCKr and the HIV envelope protein co-immunoprecipitate. Specifically, a lysate from a HIV infected cell (for example, a T-cell line infected with a HIV-1 virus strain) is incubated with recombinant BGCKr, and a anti-BGCKr antibody is used to immunoprecipitate BGCKr. An antibody directed against the HIV envelope protein is used to determine whether the HIV envelope protein interacts 1with BGCKr.

Alternatively, a cell surface binding assay can be used to determine whether BGCKr interacts with gp120 (see for example, Frankie et al. (1990) *Science*, 350:123-125; Hebert et al. (1993) *J. Biol. Chem.*, 268:18549-18553). Specifically, radioiodinated gp120 is added to cells expressing BGCKr. Specific binding of gp120 to BGCKr is tested by increasing the concentration of unlabelled gp120 and determining whether unlabelled gp120 can displace labelled gp120. Data can be analyzed by the LIGAND program (see, e.g., Huang et al.(1996) *Nat. Med.* 2:1241-1243). Binding experiments using BGCKr positive, CD4 negative cells can be carried out to determine whether the presence of soluble CD4 enhances gp120 binding.

In addition, co-receptor analysis can be carried out to determine whether BGCKr acts as a co-receptor with CD4, facilitating the entry of macrophage-tropic or T cell-tropic strains of HIV into cells (Nibbs et al. (1996) *J. Biol. Chem.* 272:32078-32083; Feng et al. (1996) *Science* 272:872-877). Briefly, a stable transfected cell line is first generated by infecting cells with recombinant vaccinia expressing cell surface CD4 and cell surface BGCKr. These cells are then infected with a luciferase reporter virus pseudotyped with envelope proteins of the macrophage-tropic HIV-1 or T-cell line adapted HIV-1. The infectivity of each HIV strain is measured four days later by assaying luciferase activity.

(see, e.g., Nibbs et al.(1997) *J. Biol. Chem.*, 32078-32083, 1997, Deng et al. (1996) *Nature* 381:661-66; Liu et al. (1996) *Cell* 86:367-77).

If appropriate, the BGCKr region(s) involved in mediating HIV entry into the cell can be determined by studying BGCKr homologs that support HIV entry and those that are unable to support HIV entry. Comparison of the sequences of the various BGCKr homologs can reveal the HIV interacting regions within BGCKr. In addition, chimeric receptors composed of a HIV co-receptor and various portions of BGCKr can be used in the co-receptor analysis assay as described above, to determine regions of BGCKr that can support HIV entry.

BGCKr appears to be a member of the G-protein family of receptors. Such receptors are capable of initiating a signalling cascade following binding to its appropriate ligand. The signalling capacity of BGCKr can be tested using a Ca^{2+} mobilization assay (Doranz et al, *J. Virul.* (1997) 71(9):6305-14). To analyze if receptor signalling affects or is required for viral entry, a mutant BGCKr which is unable to support signalling is transiently transfected into a cell line. Another cell line expressing luciferase and a HIV envelope protein is incubated with the cell line expressing nonfunctional BGCKr. Cell-cell fusion events are quantitated using the luciferase system (Doranz et al., supra).

BGCKr polymorphisms can be analyzed in high risk seronegative and seropositive HIV subjects. The BGCKr polymorphisms are detected by extracting genomic DNA from the two subject types and amplifying the BGCKr gene by PCR using primers based on the BGCKr sequence. The amplified PCR products are analyzed using gel electrophoresis and polymorphisms are detected by a change in the size of the PCR product.

Example 6: Expression analysis of BGCKr

The expression of BGCKr in various tissue samples isolated from normal Balb/C mice was measured using TaqMan quantitative PCR. The results of this study are presented in Figure 7. In this figure BGCKr expression is expressed relative to BGCKr expression in Th2 cells that have been activated by exposure (4 hours) to anti-CD3 antibody. Expression was measured in freshly isolated spleen, lymph node, bone marrow, thymus, heart, brain, lung, liver, and colon. In this study expression was highest in lymph node, bone marrow, thymus, and lung.

In a second study, expression of BGCKr in sorted cells isolated from Balb/C mice was measured using TaqMan quantitative PCR. The results of this study are presented in Figure 8. In this figure BGCKr expression is expressed relative to BGCKr expression in Th2 cells. Expression was measured in Th1 cells, Th1 cells exposed to anti-CD3 antibody for 4 hours (Th1 4h), Th2 cells, Th2 cells exposed to anti-CD3 antibody for 4 hours (Th2 4h), resting splenocytes, splenocytes activated with anti-CD3 antibody, resting CD4 cells, CD4 cells activated with anti-CD3 antibody, resting CD8 cells, CD8 cells activated with anti-CD3 antibody, resting B cells, B cells activated with LPS, peritoneal eosinophils, macrophages isolated from the peritoneal cavity after thioglycolate treatment (macrophage-TGC), and neutrophils isolated from the peritoneal cavity after thioglycolate treatment (neutrophils-TGC). Highest expression was observed in resting splenocytes, resting CD4 cells, resting CD8 cells, and resting B cells.

In a third study, the expression of BGCKr in various human tissue samples, human cell lines and isolated human cells was measured using TaqMan quantitative PCR. The results of this study are presented in Figure 9. In this figure BGCKr expression

is expressed relative to BGCKr expression in Th0 cells. Expression was measured in the following normal organ samples: lymph node, thymus, heart, lung, spleen, muscle, fetal liver, and liver. BGCKr expression was also measured in a liver fibrosis sample. In addition, peripheral blood lymphocytes (PBL) were isolated from volunteers and cultured under standard conditions to produce Th0, Th1, and Th2 cell cultures. BGCKr expression in the Th0, TH1, and Th2 cells so produced was measured in the absence of exposure to anti-CD3 antibody and after exposure to anti-CD3 antibody for 6 hours or 24 hours. BGCKr expression was also measured in various cells types isolated from PBL samples taken from volunteers. The cell populations in which BGCKr expression was measured were: naive CD3 cells, CD cells activated with anti-CD3 antibody (CD3 activated), CD4 cells, CD8 cells, resting B cells, B cells activated with LPS (B cell activated), resting PBMC, PBMC activated with PHA (PBMC activated), CD14 cells, CD16 negative cells, and granulocytes. BGCKr expression was also measured in the following human cell lines: MOLT4, K562, HL 60, 293, and HPK. BGCKr expression was also measured in Hep3b cells grown under normal oxygen tension (normaxia), Hep3b cells grown under reduced oxygen tension (hypoxia) CD34 cells isolated from a donor that had been treated with GCSF (mPB CD34+), CD34 cells isolated from adult bone marrow (ABM CD34+), and leukocytes isolated from bone marrow (BM leukocytes).

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such

equivalents are intended to be encompassed by the following claims.

What is claimed is:

1. An isolated nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence which is at least 55% identical to the nucleotide sequence of SEQ ID NO:4 or SEQ ID NO:6, the cDNA insert of the plasmid deposited with ATCC as Accession Number 209756 or a complement thereof;

b) a nucleic acid molecule comprising a fragment of at least 300 nucleotides of the nucleotide sequence of SEQ ID NO:4 or SEQ ID NO:6, the cDNA insert of the plasmid deposited with ATCC as Accession Number 209756, or a complement thereof;

c) nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:5 or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 209756;

d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2 wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:5 or the polypeptide encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 209756; and

e) a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:5 or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 209756, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NO:4 or SEQ ID NO:6 under stringent conditions.

2. The isolated nucleic acid molecule of claim 1, which is selected from the group consisting of:

a) a nucleic acid comprising the nucleotide sequence of SEQ ID NO:4 or SEQ ID NO:6, or the cDNA insert of the plasmid deposited with ATCC as Accession Number 209756, or a complement thereof; and

b) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:5 or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 209756.

3. The nucleic acid molecule of claim 1 further comprising vector nucleic acid sequences.

4. The nucleic acid molecule of claim 1 further comprising nucleic acid sequences encoding a heterologous polypeptide.

5. A host cell which contains the nucleic acid molecule of claim 1.

6. The host cell of claim 4 which is a mammalian host cell.

7. A non-human mammalian host cell containing the nucleic acid molecule of claim 1.

8. An isolated polypeptide selected from the group consisting of:

a) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:5, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:5.

b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:5 or an amino acid sequence encoded by the cDNA insert

of the plasmid deposited with ATCC as Accession Number 209756, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:4 or SEQ ID NO:6 under stringent conditions;

c) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 55% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:4 or SEQ ID NO:6.

9. The isolated polypeptide of claim 8 comprising the amino acid sequence of SEQ ID NO:5 or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 209756.

10. The polypeptide of claim 8 further comprising heterologous amino acid sequences.

11. An antibody which selectively binds to a polypeptide of claim 8.

12. A method for producing a polypeptide selected from the group consisting of:

a) a polypeptide comprising the amino acid sequence of SEQ ID NO:5 or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 209756;

b) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:5 or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 209756, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:5 or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 209756; and

c) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:5 or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 209756, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:4 or SEQ ID NO:6 under stringent conditions;

comprising culturing the host cell of claim 5 under conditions in which the nucleic acid molecule is expressed.

13. The isolated polypeptide of claim 8 comprising the amino acid sequence of SEQ ID NO:5 or an amino acid encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 209756.

14. A method for detecting the presence of a polypeptide of claim 8 in a sample, comprising:

a) contacting the sample with a compound which selectively binds to a polypeptide of claim 8; and

b) determining whether the compound binds to the polypeptide in the sample.

15. The method of claim 14, wherein the compound which binds to the polypeptide is an antibody.

16. A kit comprising a compound which selectively binds to a polypeptide of claim 8 and instructions for use.

17. A method for detecting the presence of a nucleic acid molecule of claim 1 in a sample, comprising the steps of:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and

b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample.

18. The method of claim 17, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

19. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of claim 1 and instructions for use.

20. A method for identifying a compound which binds to a polypeptide of claim 8 comprising the steps of:

a) contacting a polypeptide, or a cell expressing a polypeptide of claim 8 with a test compound; and

b) determining whether the polypeptide binds to the test compound.

21. The method of claim 20, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detecting of test compound/polypeptide binding;

b) detection of binding using a competition binding assay;

c) detection of binding using an assay for G-intracellular signalling activity.

22. A method for modulating the activity of a polypeptide of claim 8 comprising contacting a polypeptide or a cell expressing a polypeptide of claim 8 with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

23. A method for identifying a compound which modulates the activity of a polypeptide of claim 8, comprising:

a) contacting a polypeptide of claim 8 with a test compound; and

b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

```
GTCGACCCAC GCGTCCGA ATG AAT GGC ACT TAT GAC TAC AGT CAA TAT GAA          51
                    Met Asn Gly Thr Tyr Asp Tyr Ser Gln Tyr Glu
                      1               5                  10

CTG ATC TGT ATC AAA GAA GAT GTC AGA GAA TTT GCA AAA GTT TTC CTC          99
Leu Ile Cys Ile Lys Glu Asp Val Arg Glu Phe Ala Lys Val Phe Leu
            15                  20                  25

CCT GTA TTC CTC ACA ATA GTT TTC GTC ATT GGA CTT GCA GGC AAT TCC         147
Pro Val Phe Leu Thr Ile Val Phe Val Ile Gly Leu Ala Gly Asn Ser
        30                  35                  40

ATG GTA GTG GCA ATT TAT GCC TAT TAC AAG AAA CAG AGA ACC AAA ACA         195
Met Val Val Ala Ile Tyr Ala Tyr Tyr Lys Lys Gln Arg Thr Lys Thr
    45                  50                  55

GAT GTG TAC ATC CTG AAT TTG GCT GTA GCA GAT TTA CTC CTT CTA TTC         243
Asp Val Tyr Ile Leu Asn Leu Ala Val Ala Asp Leu Leu Leu Leu Phe
60                  65                  70                  75

ACT CTG CCT TTT TGG GCT GTT AAT GCA GTT CAT GGG TGG GTT TTA GGG         291
Thr Leu Pro Phe Trp Ala Val Asn Ala Val His Gly Trp Val Leu Gly
                80                  85                  90

AAA ATA ATG TGC AAA ATA ACT TCA GCC TTG TAC ACA CTA AAC TTT GTC         339
Lys Ile Met Cys Lys Ile Thr Ser Ala Leu Tyr Thr Leu Asn Phe Val
            95                  100                 105

TCT GGA ATG CAG TTT CTG GCT TGT ATC AGC ATA GAC AGA TAT GTG GCA         387
Ser Gly Met Gln Phe Leu Ala Cys Ile Ser Ile Asp Arg Tyr Val Ala
        110                 115                 120

GTA ACT AAA GTC CCC AGC CAA TCA GGA GTG GGA AAA CCA TGC TGG ATC         435
Val Thr Lys Val Pro Ser Gln Ser Gly Val Gly Lys Pro Cys Trp Ile
    125                 130                 135

ATC TGT TTC TGT GTC TGG ATG GCT GCC ATC TTG CTG AGC ATA CCC CAG         483
Ile Cys Phe Cys Val Trp Met Ala Ala Ile Leu Leu Ser Ile Pro Gln
140                 145                 150                 155

CTG GTT TTT TAT ACA GTA AAT GAC AAT GCT AGG TGC ATT CCC ATT TTC         531
Leu Val Phe Tyr Thr Val Asn Asp Asn Ala Arg Cys Ile Pro Ile Phe
                160                 165                 170

CCC CGC TAC CTA GGA ACA TCA ATG AAA GCA TTG ATT CAA ATG CTA GAG         579
Pro Arg Tyr Leu Gly Thr Ser Met Lys Ala Leu Ile Gln Met Leu Glu
            175                 180                 185

ATC TGC ATT GGA TTT GTA GTA CCC TTT CTT ATT ATG GGG GTG TGC TAC         627
Ile Cys Ile Gly Phe Val Val Pro Phe Leu Ile Met Gly Val Cys Tyr
        190                 195                 200

TTT ATC ACA GCA AGG ACA CTC ATG AAG ATG CCA AAC ATT AAA ATA TCT         675
Phe Ile Thr Ala Arg Thr Leu Met Lys Met Pro Asn Ile Lys Ile Ser
    205                 210                 215
```

FIG. 1A

```
CGA CCC CTA AAA GTT CTG CTC ACA GTC GTT ATA GTT TTC ATT GTC ACT      723
Arg Pro Leu Lys Val Leu Leu Thr Val Val Ile Val Phe Ile Val Thr
220                 225                 230                 235

CAA CTG CCT TAT AAC ATT GTC AAG TTC TGC CGA GCC ATA GAC ATC ATC      771
Gln Leu Pro Tyr Asn Ile Val Lys Phe Cys Arg Ala Ile Asp Ile Ile
                240                 245                 250

TAC TCC CTG ATC ACC AGC TGC AAC ATG AGC AAA CGC ATG GAC ATC GCC      819
Tyr Ser Leu Ile Thr Ser Cys Asn Met Ser Lys Arg Met Asp Ile Ala
            255                 260                 265

ATC CAA GTC ACA GAA AGC ATC GCA CTC TTT CAC AGC TGC CTC AAC CCA      867
Ile Gln Val Thr Glu Ser Ile Ala Leu Phe His Ser Cys Leu Asn Pro
        270                 275                 280

ATC CTT TAT GTT TTT ATG GGA GCA TCT TTC AAA AAC TAC GTT ATG AAA      915
Ile Leu Tyr Val Phe Met Gly Ala Ser Phe Lys Asn Tyr Val Met Lys
    285                 290                 295

GTG GCC AAG AAA TAT GGG TCC TGG AGA AGA CAG AGA CAA AGT GTG GAG      963
Val Ala Lys Lys Tyr Gly Ser Trp Arg Arg Gln Arg Gln Ser Val Glu
300                 305                 310                 315

GAG TTT CCT TTT GAT TCT GAG GGT CCT ACA GAG CCA ACC AGT ACT TTT     1011
Glu Phe Pro Phe Asp Ser Glu Gly Pro Thr Glu Pro Thr Ser Thr Phe
                320                 325                 330

AGC ATT TAAAGGTAAA ACTGCTCTGC CTTTTGCTTG GATACATATG AATGATGCTT     1067
Ser Ile

TCCCCTCAAA TAAAACATCT GCATTATTCT GAAACTCAAA AAAAAAAAAA AAAGGGCGGC  1127
CGC                                                                1130
```

FIG. 1B

```
                                                                M  A        2
TTAGGTCATTTGATTTTATACTCTGTACTCAAGACTGCTCCTCTCTGCCGACTACAACAGATTGGAGCC ATG GCT     75

 L  E  Q  N  Q  S  T  D  Y  Y  Y  E  E  N  E  M  N  G  T  Y     22
TTG GAA CAG AAC CAG TCA ACA GAT TAT TAT TAT GAG GAA AAT GAA ATG AAT GGC ACT TAT   135

 D  Y  S  Q  Y  E  L  I  C  I  K  E  D  V  R  E  F  A  K  V     42
GAC TAC AGT CAA TAT GAA CTG ATC TGT ATC AAA GAA GAT GTC AGA GAA TTT GCA AAA GTT   195

 F  L  P  V  F  L  T  I  V  F  V  I  G  L  A  G  N  S  M  V     62
TTC CTC CCT GTA TTC CTC ACA ATA GTT TTC GTC ATT GGA CTT GCA GGC AAT TCC ATG GTA   255

 V  A  I  Y  A  Y  Y  K  K  Q  R  T  K  T  D  V  Y  I  L  N     82
GTG GCA ATT TAT GCC TAT TAC AAG AAA CAG AGA ACC AAA ACA GAT GTG TAC ATC CTG AAT   315

 L  A  V  A  D  L  L  L  L  F  T  L  P  F  W  A  V  N  A  V    102
TTG GCT GTA GCA GAT TTA CTC CTT CTA TTC ACT CTG CCT TTT TGG GCT GTT AAT GCA GTT   375

 H  G  W  V  L  G  K  I  M  C  K  I  T  S  A  L  Y  T  L  N    122
CAT GGG TGG GTT TTA GGG AAA ATA ATG TGC AAA ATA ACT TCA GCC TTG TAC ACA CTA AAC   435

 F  V  S  G  M  Q  F  L  A  C  I  S  I  D  R  Y  V  A  V  T    142
TTT GTC TCT GGA ATG CAG TTT CTG GCT TGT ATC AGC ATA GAC AGA TAT GTG GCA GTA ACT   495

 K  V  P  S  Q  S  G  V  G  K  P  C  W  I  I  C  F  C  V  W    162
AAA GTC CCC AGC CAA TCA GGA GTG GGA AAA CCA TGC TGG ATC ATC TGT TTC TGT GTC TGG   555

 M  A  A  I  L  L  S  I  P  Q  L  V  F  Y  T  V  N  D  N  A    182
ATG GCT GCC ATC TTG CTG AGC ATA CCC CAG CTG GTT TTT TAT ACA GTA AAT GAC AAT GCT   615

 R  C  I  P  I  F  P  R  Y  L  G  T  S  M  K  A  L  I  Q  M    202
AGG TGC ATT CCC ATT TTC CCC CGC TAC CTA GGA ACA TCA ATG AAA GCA TTG ATT CAA ATG   675

 L  E  I  C  I  G  F  V  V  P  F  L  I  M  G  V  C  Y  F  I    222
CTA GAG ATC TGC ATT GGA TTT GTA GTA CCC TTT CTT ATT ATG GGG GTG TGC TAC TTT ATC   735
```

FIG. 2A

```
T  A  R  T  L  M  K  M  P  N  I  K  I  S  R  P  L  K  V  L   242
ACA GCA AGG ACA CTC ATG AAG ATG CCA AAC ATT AAA ATA TCT CGA CCC CTA AAA GTT CTG   795

L  T  V  V  I  V  F  I  V  T  Q  L  P  Y  N  I  V  K  F  C   262
CTC ACA GTC GTT ATA GTT TTC ATT GTC ACT CAA CTG CCT TAT AAC ATT GTC AAG TTC TGC   855

R  A  I  D  I  I  Y  S  L  I  T  S  C  N  M  S  K  R  M  D   282
CGA GCC ATA GAC ATC ATC TAC TCC CTG ATC ACC AGC TGC AAC ATG AGC AAA CGC ATG GAC   915

I  A  I  Q  V  T  E  S  I  A  L  F  H  S  C  L  N  P  I  L   302
ATC GCC ATC CAA GTC ACA GAA AGC ATC GCA CTC TTT CAC AGC TGC CTC AAC CCA ATC CTT   975

Y  V  F  M  G  A  S  F  K  N  Y  V  M  K  V  A  K  K  Y  G   322
TAT GTT TTT ATG GGA GCA TCT TTC AAA AAC TAC GTT ATG AAA GTG GCC AAG AAA TAT GGG  1035

S  W  R  R  Q  R  Q  S  V  E  E  F  P  F  D  S  E  G  P  T   342
TCC TGG AGA AGA CAG AGA CAA AGT GTG GAG GAG TTT CCT TTT GAT TCT GAG GGT CCT ACA  1095

E  P  T  S  T  F  S  I  *                                     351
GAG CCA ACC AGT ACT TTT AGC ATT TAA                           1122

AGGTAAAACTGCTCTGCCTTTTGCTTGGATACATATGAATGATGCTTTCCCCTCAAATAAAACATCTGCATTATTCTGA  1201

AACTCAAAAAAAAAAAAAAAAAGGGCGGCCGC                              1232
```

FIG. 2B

```
 M  A  L  E  L  N  Q  S  A  E  Y  Y  Y  E  E  N  E  M  N  Y      20
ATG GCT CTG GAG CTG AAC CAG TCA GCG GAA TAC TAT TAT GAG GAG AAT GAG ATG AAC TAC    60

 T  H  D  Y  S  Q  Y  E  V  I  C  I  K  E  E  V  R  Q  F  A      40
ACT CAC GAC TAC AGC CAG TAC GAA GTG ATC TGC ATA AAA GAA GAG GTC AGG CAG TTC GCA   120

 K  V  F  L  P  A  F  F  T  V  A  F  V  T  G  L  A  G  N  S      60
AAA GTC TTC CTG CCC GCC TTC TTC ACA GTG GCC TTT GTC ACT GGG CTC GCA GGG AAC TCC   180

 V  V  V  A  I  Y  A  Y  Y  K  K  Q  R  T  K  T  D  V  Y  I      80
GTA GTT GTG GCG ATT TAC GCC TAT TAC AAG AAG·CAG AGG ACC AAG ACC GAT GTG TAC ATC   240

 L  N  L  A  V  A  D  L  L  L  L  I  T  L  P  F  W  A  V  N     100
CTG AAC CTG GCT GTA GCA GAC TTG TTA CTT CTG ATC ACG CTG CCT TTC TGG GCA GTT AAT   300

 A  V  H  G  W  I  L  G  K  M  M  C  K  V  T  S  A  L  Y  T     120
GCA GTT CAC GGA TGG ATT CTA GGC AAA ATG ATG TGC AAA GTA ACG TCA GCC CTG TAC ACG   360

 V  N  F  V  S  G  M  Q  F  L  A  C  I  S  I  D  R  Y  W  A     140
GTA AAC TTT GTC TCT GGG ATG CAG TTC CTG GCT TGT ATC AGC ATT GAC AGA TAT TGG GCA   420

 I  T  K  A  P  S  Q  S  G  A  G  R  P  C  W  I  I  C  C  C     160
ATT ACC AAA GCC CCC AGC CAA TCA GGA GCG GGG AGA CCC TGT TGG ATC ATC TGT TGC TGT   480

 V  W  M  A  A  I  L  L  S  X  P  Q  L  V  F  Y  T  V  N  Q     180
GTG TGG ATG GCC GCC ATC TTG CTG AGC ATN CCC CAG CTG GTT TTT TAC ACA GTG AAT CAA   540

 N  A  R  C  T  P  I  F  P  H  H  L  G  T  S  L  K  A  S  I     200
AAT GCT AGG TGC ACT CCC ATC TTT CCC CAC CAC CTA GGA ACA TCC CTG AAA GCA TCC ATT   600

 Q  M  L  E  I  G  I  G  F  V  V  P  F  L  I  M  G  V  C  Y     220
CAG ATG CTG GAA ATC GGC ATC GGC TTT GTG GTC CCG TTT CTC ATC ATG GGC GTG TGC TAT   660

 A  S  T  A  R  A  L  I  K  M  P  N  I  K  K  S  R  P  L  R     240
GCC AGT ACC GCC AGG GCG CTC ATC AAG ATG CCC AAC ATT AAA AAG TCT CGC CCC CTC AGG   720

 V  L  L  A  V  V  V  V  F  I  V  T  Q  L  P  Y  N  V  V  K     260
GTT CTG CTC GCG GTG GTG GTG GTT TTC ATT GTC ACC CAG CTG CCC TAT AAC GTT GTT AAG   780

 F  C  Q  A  I  D  A  I  Y  L  L  I  T  S  C  D  M  S  K  R     280
TTC TGC CAA GCC ATA GAT GCC ATC TAC CTG CTG ATC ACC AGC TGC GAT ATG AGC AAA CGC   840

 M  D  V  A  I  Q  V  T  E  S  I  A  L  F  H  S  C  L  N  P     300
ATG GAT GTC GCC ATC CAA GTC ACA GAG AGC ATC GCG CTC TTC CAC AGC TGC CTC AAC CCC   900

 I  L  Y  V  F  M  G  A  S  F  K  N  Y  I  M  K  V  A  K  K     320
ATC CTG TAT GTC TTC ATG GGG GCC TCC TTC AAA AAC TAT ATC ATG AAA GTG GCC AAG AAA   960

 Y  G  S  W  R  R  Q  R  Q  N  V  E  E  I  P  F  D  S  E  G     340
TAT GGA TCC TGG AGA AGA CAG AGA CAG AAC GTG GAA GAA ATT CCT TTT GAT TCT GAG GGT  1020

 P  T  E  P  T  S  S  F  T  I  *                                351
CCT ACA GAG CCA ACC AGT TCT TTT ACC ATT TAA                     1053
```

FIG. 3

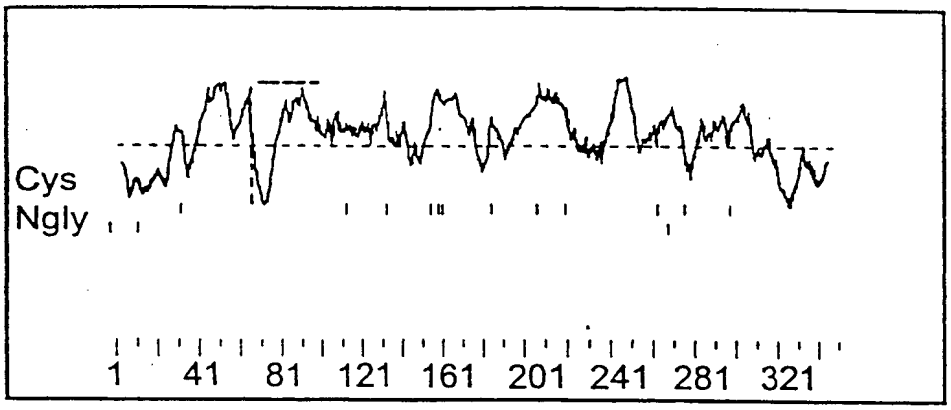


FIG. 4

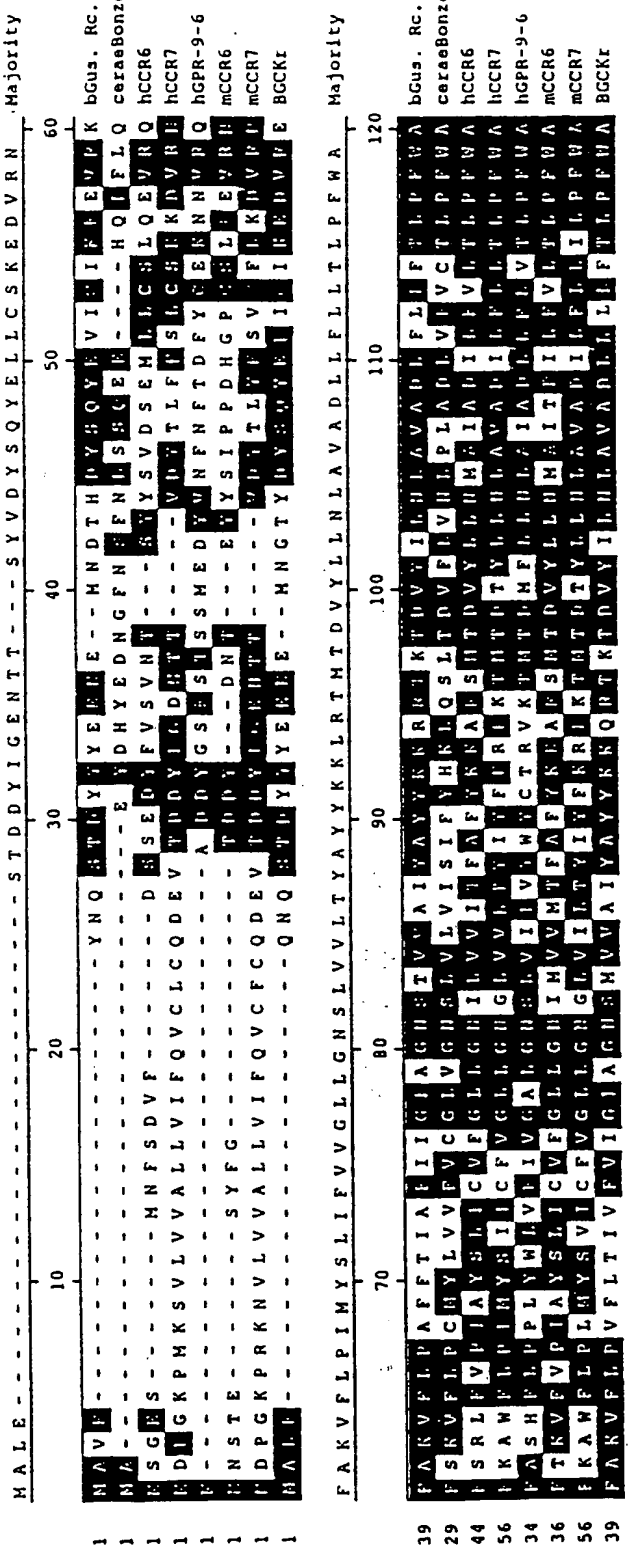

Majority
bGus. Rc.
cerasBonzo
hCCR6
hCCR7
hGPR-9-6
mCCR6
mCCR7
BGCKr

FIG. 5A

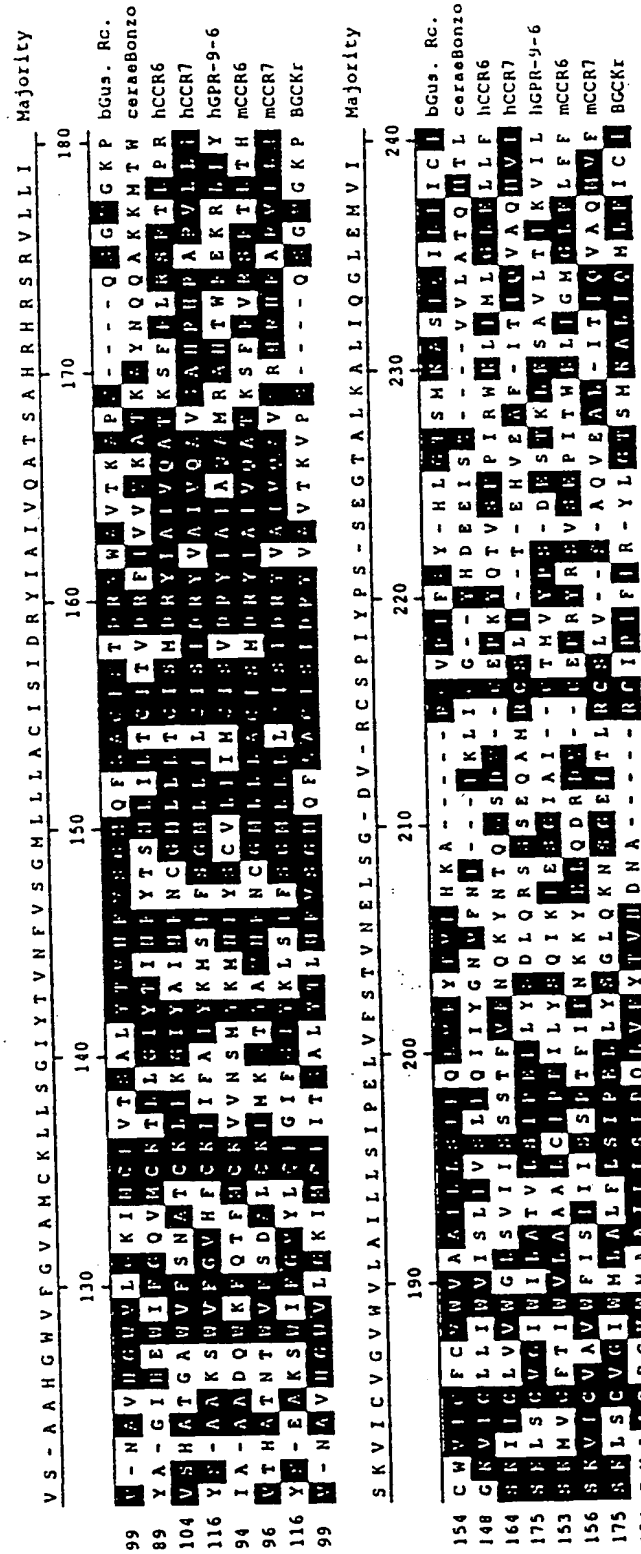

Majority
bGus. Rc.
caraeBonzo
hCCR6
hCCR7
hGPR-9-6
mCCR6
mCCR7
BGCKr

FIG. 5B

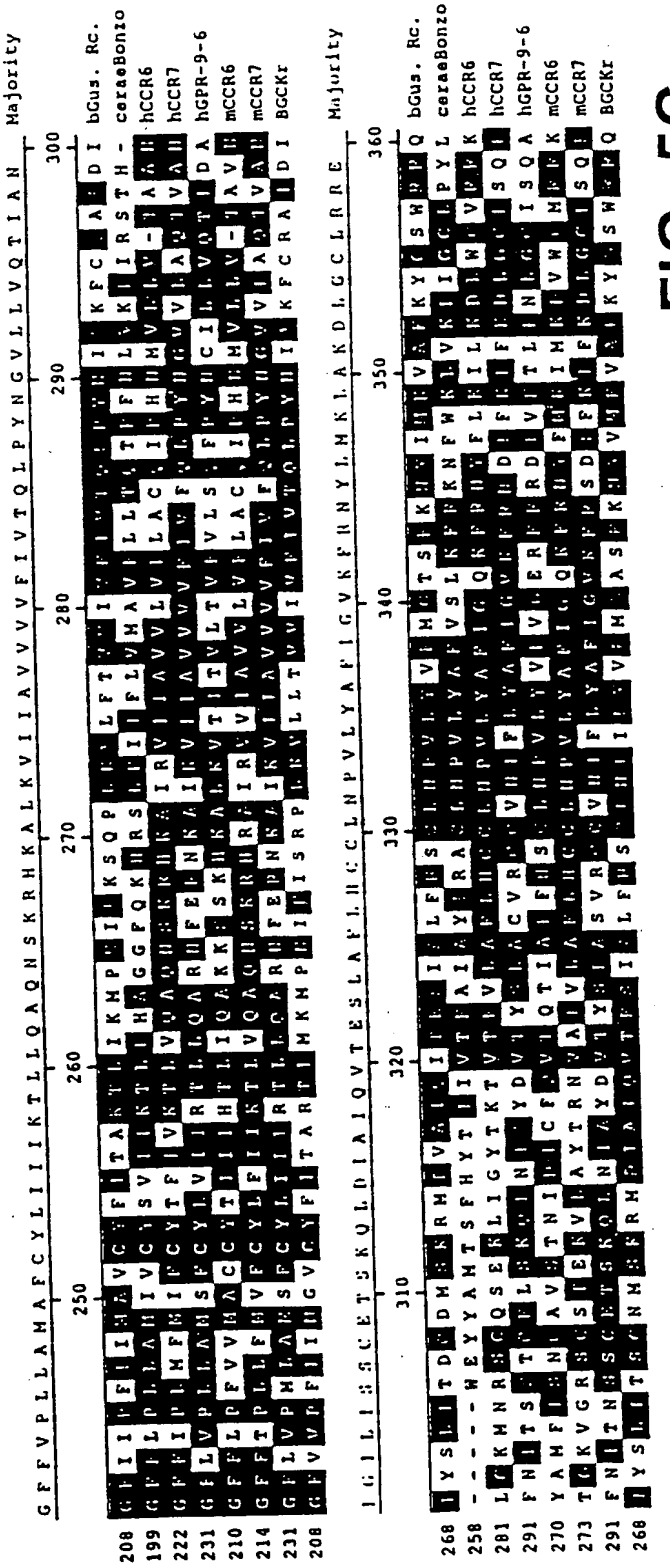
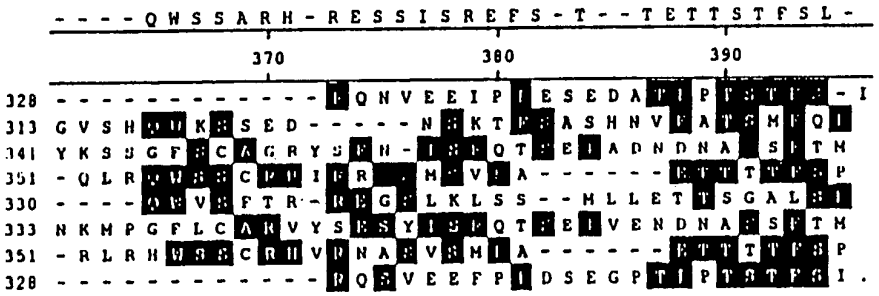


FIG. 5C

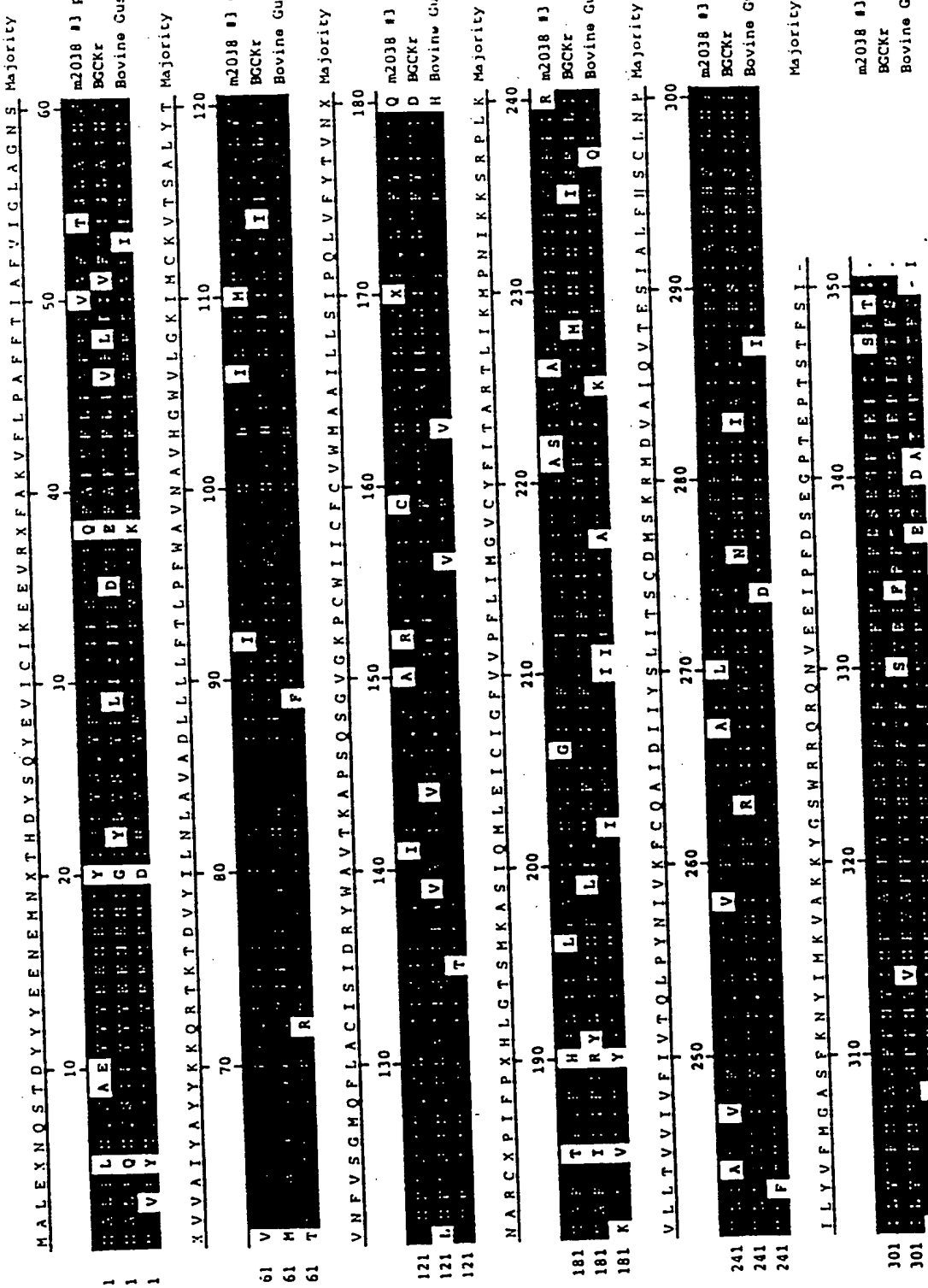


Majority

bGus. Rc.
ceraeBonzo
hCCR6
hCCR7
hGPR-9-6
mCCR6
mCCR7
BGCKr

Decoration 'Decoration #1': Shade (with solid black) residues that match the Consensus exactly.

FIG. 5D



Decoration '3': Shade (with solid black) residues that match the Consensus exactly.

FIG. 6

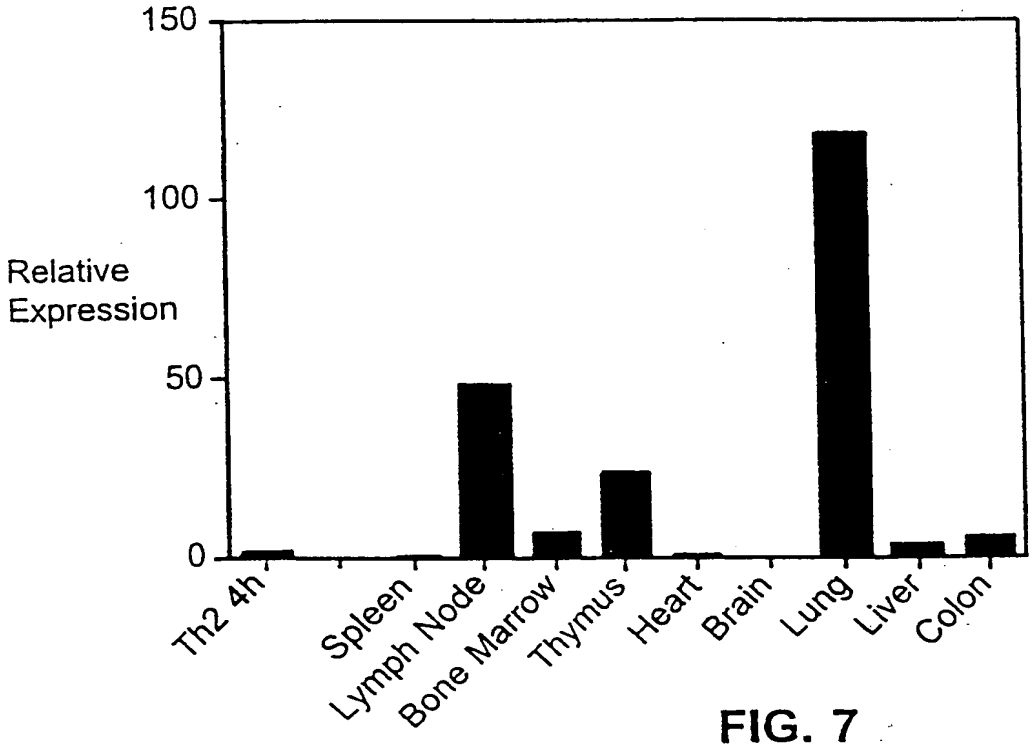

150
100
50
0
Relative Expression
Th2 4h
Spleen
Lymph Node
Bone Marrow
Thymus
Heart
Brain
Lung
Liver
Colon

FIG. 7

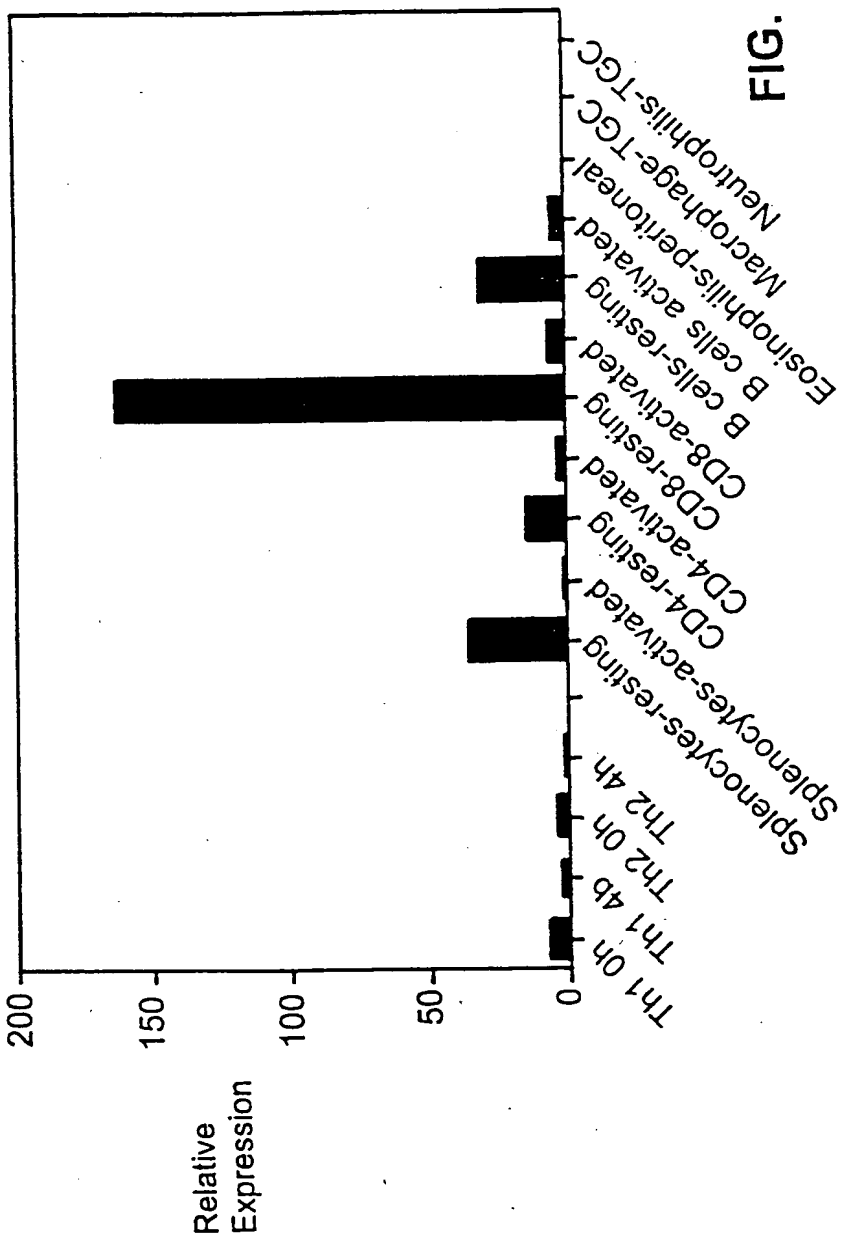

Relative Expression
200
150
100
50
0
Th1 0h
Th1 4h
Th2 0h
Th2 4h
Splenocytes-resting
Splenocytes-activated
CD4-resting
CD4-activated
CD8-resting
CD8-activated
B cells-resting
B cells-activated
Eosinophilis-peritoneal
Macrophage-TGC
Neutrophilis-TGC

FIG. 8

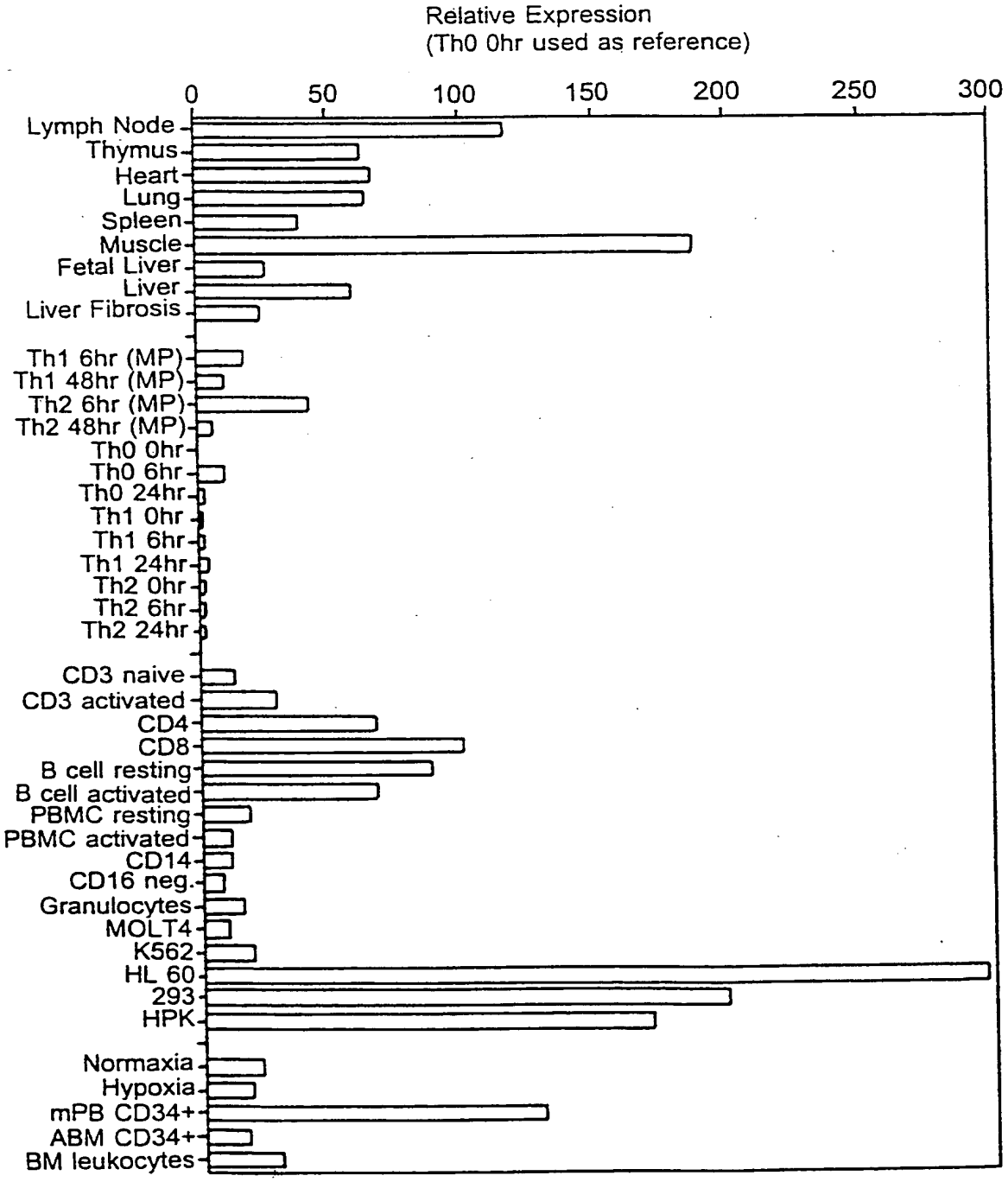

Relative Expression
(Th0 0hr used as reference)
0
50
100
150
200
250
300
Lymph Node
Thymus
Heart
Lung
Spleen
Muscle
Fetal Liver
Liver
Liver Fibrosis
Th1 6hr (MP)
Th1 48hr (MP)
Th2 6hr (MP)
Th2 48hr (MP)
Th0 0hr
Th0 6hr
Th0 24hr
Th1 0hr
Th1 6hr
Th1 24hr
Th2 0hr
Th2 6hr
Th2 24hr
CD3 naive
CD3 activated
CD4
CD8
B cell resting
B cell activated
PBMC resting
PBMC activated
CD14
CD16 neg.
Granulocytes
MOLT4
K562
HL 60
293
HPK
Normaxia
Hypoxia
mPB CD34+
ABM CD34+
BM leukocytes

FIG. 9

```
*->GNlLVilvilrtkklrtptnifilNLAvADLLflltlppwalyylvg
   GN++V+++ +++kk rt t+++ilNLAvADLL+l+tlp+wa+ +++g
58 GNSMVVAIYAYYKKQRTKTDVYILNLAVADLLLLFTLPFWAVNAVHG 104

   gsedWpfGsalCklvtaldvvnmyaSillLtaISiDRYlAIvhPlryrrr
      W++G+++Ck+++al+++n+ +++ +L++ISiDRY+A+++  + + +
105 ----WVLGKIMCKITSALYTLNFVSGMQFLACISIDRYVAVTKVPSQSGV 150

   rtsprrAkvvillvWvlalllslPpllfswvktveegngtlnvmvtvCli
         +++++ +vW +a+lls+P+l+f++v++++              C +
151 GK---PCWIICFCVWMAAILLSIPQLVFYTVNDNAR----------CIP 186

   dfpeestasvstwlrsyvllstlvgFllPllvilvcYtrIlrtlr.....
    fp      ++   ++ l + +gF++P+l++ vcY+ ++rtl + ++
187 IFPRYL---GTSMKALIQMLEICIGFVVPFLIMGVCYFITARTLMkmpni 233

   ...kaaktllvvvvvFvlCWlPyfivllldtlc.lsiimsstCelervlp
   + ++ +k+ll+vv+vF++ +lPy+iv + +++  + +  + +C +++  +
234 kisRPLKVLLTVVIVFIVTQLPYNIVKFCRAIDiIYSL-ITSCNMSKRMD 282

   tallvtlwLayvNsclNPiIY<-*
   +a++vt+ +a ++sclNPi+Y
283 IAIQVTESIALFHSCLNPILY     303
```

FIG. 10